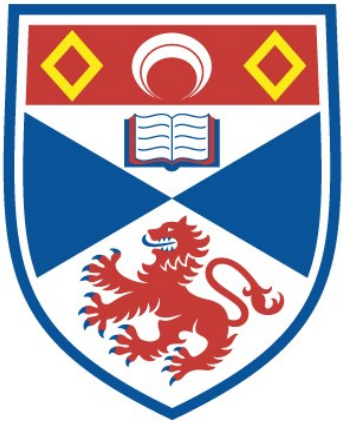
# DEVELOPMENT OF A BIOTECHNOLOGICAL TOOLKIT FOR THE SYNTHESIS OF DIVERSE CYCLIC PEPTIDES

Gregory Mann

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



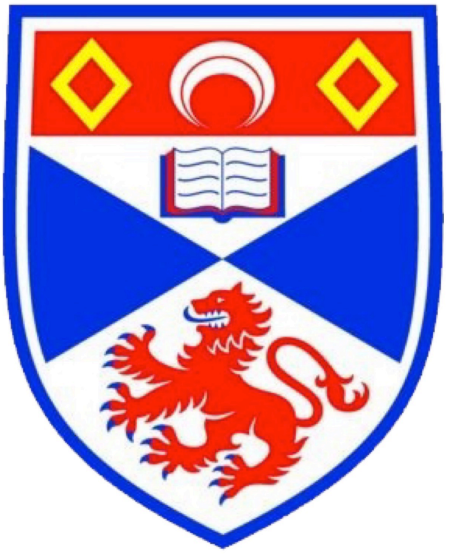
2017

Full metadata for this item is available in
St Andrews Research Repository
at:
http://research-repository.st-andrews.ac.uk/

Please use this identifier to cite or link to this item:
http://hdl.handle.net/10023/10826

This item is protected by original copyright

# Development of a Biotechnological Toolkit for the Synthesis of Diverse Cyclic Peptides



by

Gregory Mann

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy

in the

School of Chemistry

**University of St Andrews**

November 2016

# Declaration of Authorship

I, Gregory Mann, hereby certify that this thesis, which is approximately 40 000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September, 2012 and as a candidate for the degree of Doctor of Philosophy in August, 2016; the higher study for which this is a record was carried out in the University of St Andrews between 2012 and 2016.

Signature of candidate: _____

Date: _____

Signature of supervisor: _____

Date: _____

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Signature of supervisor: _____

Date: _____

# Permission for Publication

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of work may be made and supplied to an bona fide library or research worker, that my thesis will be electronically accessibly for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of the thesis:

Access to all of printed copy but embargo of all electronic publication of thesis for a period of one year on the following grounds:

Publication would be commercially damaging to the researcher, or to the supervisor, or the University.

Signature of candidate: _____

Date: _____

Signature of supervisor: _____

Date: _____

*"Iustus impetro fiat."*

– Anonymous

# *Abstract*

Cyclic peptides possess desirable characteristics as potential pharmaceutical scaffolds. The cyanobactin family of cyclic peptide natural products boast diverse structures and bioactivity. Exemplars are the patellamides, which have attracted attention due to their ability to reverse the effects of multi-drug resistance in human leukemia cells. In addition to their macrocyclic architecture patellamides contain azol(in)e heterocycles and D-amino acids. This structural complexity makes them challenging targets for chemical synthesis. Understanding their biosynthesis will enable the development of a biotechnological 'toolkit' for the synthesis of new pharmaceutical compounds. Patellamides are ribosomally-synthesised and post-translationally modified peptides (RiPPs) and much of their biosynthesis has been elucidated, however there are still elements of their biosynthesis that are not yet fully understood.

PatA and PatG contain C-terminal domains of unknown function (DUFs). The crystal structure of PatG-DUF has been solved and subsequent to biochemical and biophysical investigation PatG-DUF was found not to constitute an essential part of the biotechnological 'toolkit' and can be excluded from *in vitro* enzyme-based synthesis of cyanobactin-like cyclic peptides.

The cyanobactin heterocyclases are able to introduce heterocycles into a peptide backbone, seemingly irrespective of the neighbouring residues; however a molecular rational governing substrate recognition is unknown. Additionally the mechanism of heterocyclisaton is disputed. Analysis of crystal structures of LynD in complex with cofactor and substrate (solved by Dr Jesko Koehnke) enabled the active site and substrate recognition site to be located. A new mechanism for heterocyclisation has been proposed. Guided by the substrate recognition observed in complex structures a constituently active heterocyclase (AcLynD) has been engineered, which is able to process short, leaderless peptide substrates.

Epimerisation in cyanobactin biosynthesis is believed to be spontaneous, but its precise timing is uncertain. NMR analysis of selectively labelled peptide substrates processed by the modifying enzymes, identified epimerisation to be spontaneous on the macrocycle, regardless of whether the neighbouring heterocycles have been oxidised.

A one-pot *in vitro* synthesis of cyanobactins has been developed, and employed to create a number of patellamide D analogues to ascertain structural-activity relationships.

# *Acknowledgements*

I wish to thank my supervisor Prof. Jim Naismith for the opportunity to study for my PhD and for his support and guidance throughout. I thank Dr Jesko Koehnke for his training and tuition. Whether I care to admit it or not I am a better scientist because of his teaching. Special thanks go to Dr Andrew Bent, Dr Emilia Ouies and Ms Laura Woodward for their ongoing support and advice. Importantly they were always there when I needed to moan! I also thank Ms Laura Woodward for proofreading this thesis. I thank my project team, the affectionally named 'Team Pata Pata' (thanks of course to Dr Lucile Moyne for introducing the team name): Dr Andrew Bent, Dr Emilia Ouies, Dr Brunello Nardone, Ms Ying Ge, Mr Hannes Ludewig, Dr Clarissa Melo Czekster, Dr Haigang Song and all the project students, who have helped provide an invaluable collaborative environment necessary for the success of this project. I wish to thank the entire JHN lab and level 2 of the BSRC annex. A special mention to Mr Robert Pengelly, Ms Laura Griffin, Dr Audrey Le Bas, Ms Luana Ferrara, Ms Miriam Weckener and Dr David Owen. Special thanks go to Dr Catherine Botting and Dr Sally Shirran for their efficient mass specrometry service - I daren't count the number of samples they have run for me over the years! Thanks to Ian Armitt for all his hard work in the media kitchen. The department always noticed when he was away. Thanks to Dr Tomas Lebl, Mrs Melanja Smith and especially Dr Uli Schwarz-Linek for their help with NMR analysis. Thank you to Prof. Nick Westwood for his useful discussion, particularly regarding analyses of final products.

I wish to thank my parents for their encouragement throughout, and for repeatedly asking me "What is it you study again?" and then simply grinning during my response before eventually telling me how clever I am. Finally I wish to thank my girlfriend Tina, for her encouragement and patience. I thank her for waiting for me to finish my PhD, and for not running away back to Germany, especially given the 'Brexit' debacle.

Oh, and of course I thank you - my examiners, for reading this thesis. I hope you had as much fun reading it as I did writing it...

# Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| ADP | adenosine diphosphate |
| Ala (A) | alanine |
| AMP | adenosine monophosphate |
| AMPCPP | α-β-methylene adenosine triphosphate |
| AMPPCP | β-γ-methylene adenosine triphosphate |
| AMPPNP | β-γ-imido adenosine triphosphate |
| Arg (R) | arginine |
| ArtGox | oxidase from athrospiramide pathway |
| Asn (N) | asparagine |
| Asp (D) | aspartic acid |
| ATP | adenosine triphosphate |
| Bicine | N,N-bis(2-hydroxyethyl)glycine |
| Bis-tris | bis-(2-hydroxyethyl)amino-tris(hydroxymethyl)methane |
| CD | circular dichroism |
| CsA | cyclosporin A |
| $C^{ThH}$ | thiazoline |
| $C^{Thz}$ | thiazole |

| | |
|---|---|
| CyaGox | oxidase from cyanothecamide pathway |
| Cys (C) | cysteine |
| DTT | 1,4 dithiothreitol |
| DUF | domain of unknown function |
| ESI | electrospray ionisation |
| FMN | flavin mononucleotide |
| Gln (Q) | glutamine |
| Glu (E) | glutamic acid |
| Gly (G) | glycine |
| HEPES | 4-(2-hydroxyethly)-1-pipazineethanesulfonic acid |
| His (H) | histidine |
| HPLC | high performance liquid chromatography |
| HSQC | heteronuclear single quantum coherence |
| Ile (I) | isoleucine |
| IPTG | isopropyl-β-D-thiogalactopyranoside |
| ITC | isothermal titration calorimetry |
| LAP | linear azole-containing peptide |
| LB | Luria Bertani |
| LC | liquid chromatography |
| Leu (L) | leucine |
| Lys (K) | lysine |
| MALDI | matrix assisted laser adsorption ionisation |

| | |
|---|---|
| Met (M) | methionine |
| MDR | multi drug resistance |
| MicD | heterocyclase from microcyclamide pathway |
| MR | molecular replacement |
| MS | mass spectrometry |
| mW | molecular weight |
| MWD | multi wavelength detector |
| n-BuOH | n-butanol |
| NMR | nuclear magnetic resonance |
| NRPS | non ribosomal peptide synthetase |
| Pgp | p-glycoprotein |
| Phe (F) | phenylalanine |
| Pi | inorganic phosphate |
| PPi | pyrophosphate |
| PPI | protein-protein interaction |
| Pro (P) | proline |
| PRPS | post-ribosomal peptide synthesis |
| PTM | post translational modification |
| qNMR | quantitative nuclear magnetic resonance |
| RiPP | ribosomally synthesised and post-translationally modified peptide |
| RMSD | root-mean square deviation |
| RP-HPLC | reverse phase high performance liquid chromatography |

| | |
|---|---|
| SAD | single wavelength anomalous diffraction |
| SDS-PAGE | sodium dodecyl sulfate - polyacrylamide gel electrophoresis |
| SeMet | selenomethionine |
| Ser (S) | serine |
| SUMO | small ubiquitin like modifier |
| TCEP | tris(2-carboxyethyl)phosphine |
| TEV | tobacco etch virus |
| Thr (T) | threonine |
| $T^{MeOx}$ | methyl oxazoline |
| $T^{MeOxz}$ | methyl ozazole |
| TOF | time of flight |
| $T_R$ | retention time |
| Tris | tris-(hydroxymethyl) aminomethane |
| Trp (W) | tryptophan |
| TruD | heterocyclase from trunkamide pathway |
| Tyr (Y) | tyrosine |
| UV | ultra violet |
| Val (V) | valine |

# Chapter 1

# Introduction

## 1.1 Cyclic Peptides as Pharmaceuticals

Nature harbours a diverse array of bioactive cyclic peptides, many of which possess desirable characteristics and are exploited as therapeutic agents[1–6] (Fig. 1.1). The examples in Fig. 1.1 demonstrate the variability in size and chemical composition in cyclic peptides offered by nature, as well as the diversity of their clinical applications.

As potential pharmaceuticals, peptides exhibit a number of benefits over more traditional small molecule compounds[7–9]. Their increased size and complexity dramatically increases their specificity towards their biological targets, thus reducing undesired side effects — a common drawback of small molecule drug candidates[7,8]. In addition they are able to disrupt protein-protein interactions (PPIs). PPIs typically involve large surface areas lacking defined binding pockets, and are therefore more difficult to target with small molecules, but are more amenable to interactions with peptides[8,10,11]. Underpinned by an enhanced knowledge and understanding of the molecular mechanisms governing different disease pathways, the inhibition or stabilisation of specific PPIs between subunits of essential multi-protein complexes are increasingly arising as promising alternative pharmaceutical targets, highlighting the importance of this advantage[12].

Cyclic peptides are of particular interest as they offer pharmacological advantages over their linear counterparts[13,14]: A decrease in conformational flexibility of a macrocycle results in a lower entropic penalty upon binding, thus increasing binding

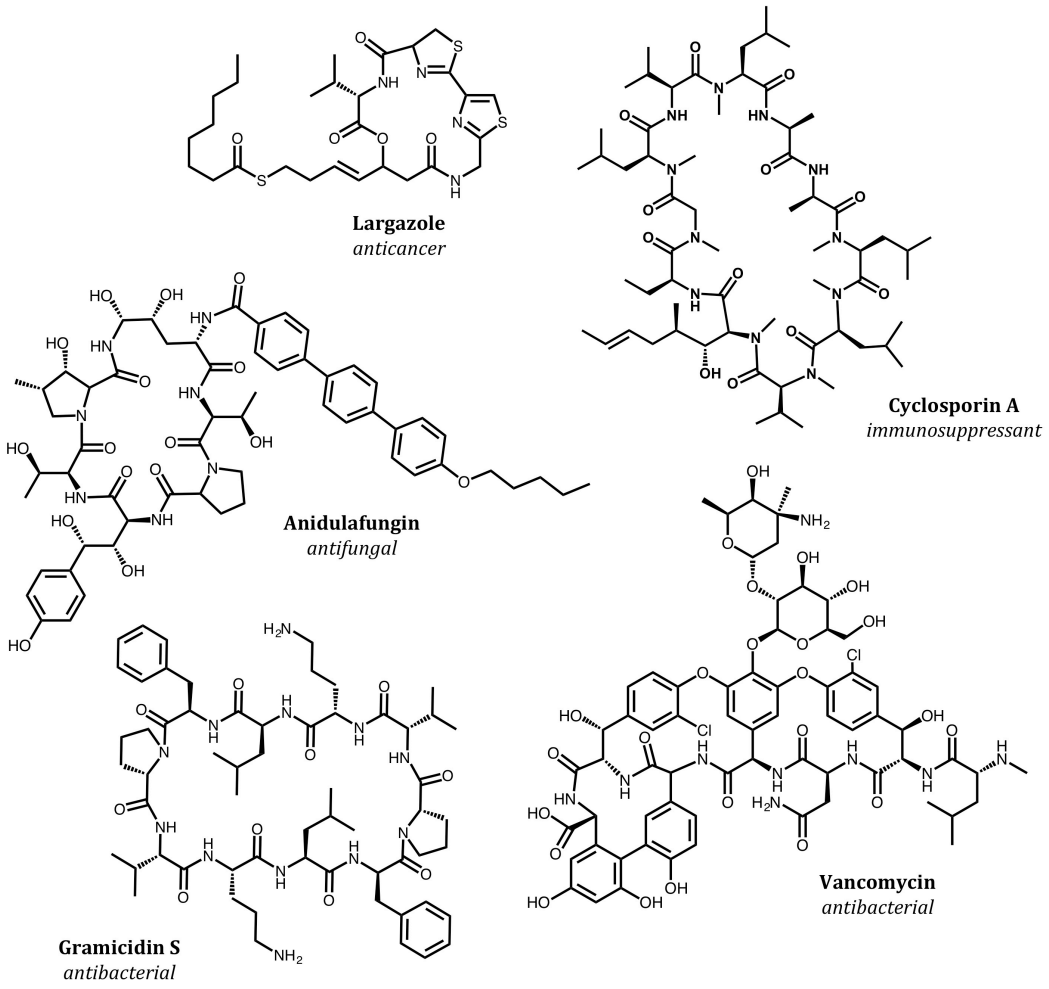



FIGURE 1.1: **Clinically available cyclic peptide drugs** Examples of structurally diverse, clinically available cyclic peptide drugs.

affinity[14]. Furthermore the absence of ionized termini facilitate easier passage across lipid membranes[15,16], and provide resistance to exo-protease degradation[11]. These advantageous properties of cyclic peptides go some way to address the shortcomings of peptides compared with small molecules as potential drug molecules[9].

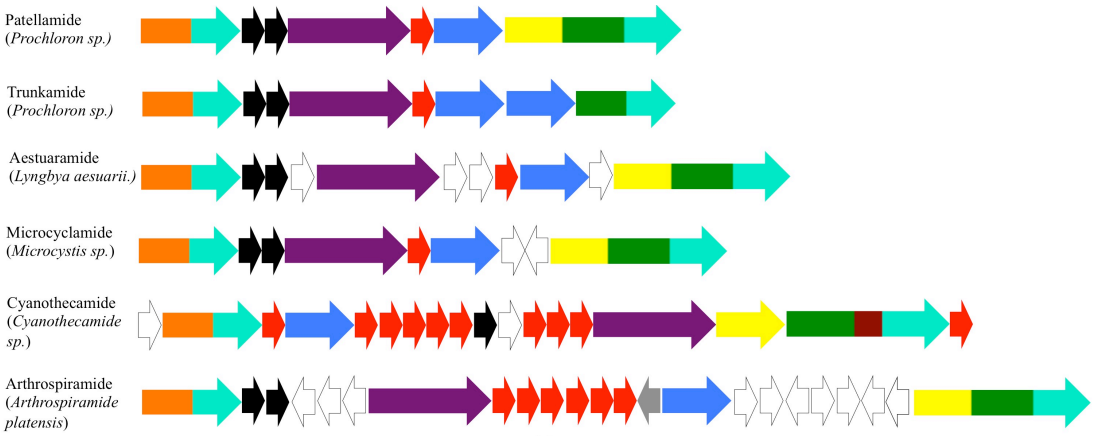
Given these desirable characteristics, the engineering of existing, and the design of novel cyclic peptides, is an excellent route to new pharmaceutical compounds. However, like other natural products, their complex architecture means that their chemical synthesis is challenging and is often expensive and low yielding[6,17]. As a consequence, there is significant interest in studying the biosynthesis of cyclic peptide natural products, with the aim to utilise the efficient and green chemistry of enzymes to generate a diverse range of novel pharmaceutical compounds.

## 1.2 Biosynthesis of Cyclic Peptides

Originally it was believed that most cyclic peptide (as well as modified linear peptide) natural products were the products of Non-Ribosomal Peptide Synthetases (NRPSs), large modular protein complexes, where each module, through its associated domains, is responsible for the incorporation (and subsequent modification if required) of a specific amino acid into a growing peptide chain[18]. However, since the turn of the century a growing number of highly modified cyclic and linear peptides have been found to be products of Post-Ribosomal Peptide Synthesis (PRPS) pathways[19,20]. These ribosomally-synthesised and post-translationally modified peptides (RiPPs) are synthesised initially by the ribosome as a longer precursor peptide, which undergoes extensive enzymatic post-translational modification (PTM) to become the final product[19–21]. Although the ribosome is canonically limited to the incorporation of proteinogenic amino acids into a peptide, the numerous tailoring enzymes from various PRPS pathways are capable of a wide range of modifications including, proteolysis of the precursor peptide, dehydration, cyclodehydration (heterocyclisation), prenylation, methylation and macrocyclisation amongst others, resulting in RiPPs displaying vast chemical and structural variation, as well as a broad array of biological activities[19,20,22].

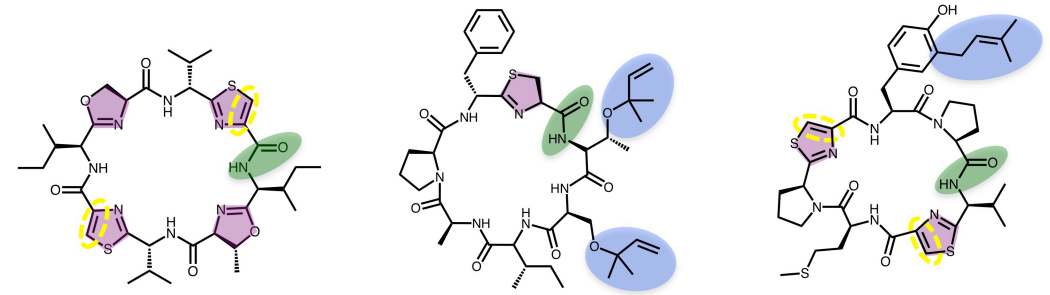
Exemplars of the RiPP family of natural products are the cyanobactins. Some examples of cyanobactins, and the biosynthetic gene clusters and precursor peptides from which they originate can be seen in Fig. 1.2. Cyanobactins, secondary metabolites of both free-living and symbiotic cyanobacteria, are a class of cyclic, and to a lesser extent, linear peptides, three to twenty amino acids in size, boasting a wealth of structural variation and promising bioactivities[21]. Patellamides A, B and C have been reported to possesses moderate antitumour activity, displaying $IC_{50}$ values in the low μg $mL^{-1}$ range against a number of human cancer cell lines[23]. Moreover, patellamides B, C and D have been shown to reduce the effects of multidrug resistance (MDR) on leukemic cell lines[24,25]. The $IC_{50}$ value for vinblastine against drug resistant CCRF-CEM human leukemic cells decreases from 90 mM to 12 mM in the presence of 2.5 μg $mL^{-1}$ patellamide B and C[16,24]. Similarly the addition of 3.3 μM patellamide D to the multidrug resistant CEM/VLB100 cell line causes the $IC_{50}$ value of vinblastine, colchicine and adriamycin to fall from 100 ng $mL^{-1}$, 140 ng $mL^{-1}$ and more than 1000 ng $mL^{-1}$ to 1.5 ng $mL^{-1}$, 50 ng $mL^{-1}$ and 110 ng $mL^{-1}$ respectively[16,25]. Encouraging antitumour activity has been

reported for trunkamide A, displaying $IC_{50}$ = 0.5 μg $mL^{-1}$ against HT-29 human colon carinoma and $IC_{50}$ = 1 μg $mL^{-1}$ against MEL-28 human melanoma cell lines[26]; and for ulithiacyclamide with $IC_{50}$ = 0.35 μg $mL^{-1}$ and $IC_{50}$ = 35 ng $mL^{-1}$ against L210 murine leukemia and KB cell lines respectively[26]. Antimalarial activity has been reported for the aerucyclamide family, with aerucyclamides B, C and D displaying $IC_{50}$ = 0.7, 2.3 and 6.3 μM respectively[26].




(a)


(b)




(c)

FIGURE 1.2: **Cyanobactins** (a) Gene clusters from patellamide (*pat*), trunkamide (*tru*), aestuaramide (*lyn*), microcyclamide (*mic*), cyanothecamide (*cya*) and arthrospiramide (*art*) pathways. Gene clusters encode a N-terminal protease (A, orange), heterocyclase (D, purple), precursor peptide (E, red), prenyl-transferase (F, blue), oxidase (yellow), macrocyclase (G, green), two domains of unknown function (turquoise), methyl-transferase (brown) associated proteins of unknown function (black), transposase (grey) and hypothetical open reading frames (white). (b) Sequence alignment of patellamide, trunkamide, aestuaramide, microcyclamide and cyanothecamide precursor peptides. The core peptides (which are modified to become the final products) are shown in red. Recognition elements for heterocyclisation, N-terminal proteolysis and C-terminal proteolysis and macrocyclisation are shown in purple, orange and green respectively. (c) examples of natural products from the patellamide, trunkamide, aestruaramide, microcyclamide and cyanothecamide pathways. The post-translational modifications are highlighted with the same colour as the enzymes that perform the transformation.

The details of RiPP biosynthesis, highlighting characteristics which make PRPS systems promising targets for bioengineering with the intention of pharmaceutical development, will be discussed using cyanobactins as an example, with particular focus on patellamide biosynthesis[21,27,28]. Cyanobactin biosynthesis is exemplified by the patellamide case, with many pathways sharing the same arrangement of the biosynthetic genes, high homology between the biosynthetic enzymes, and even significant conservation of the leader peptide and enzyme recognition sequences within the precursor peptides[21] (Fig. 1.2). Consequently the subsequent discussion regarding the enzymatic details of patellamide biosynthesis can be thought of as broadly applicable to cyanobactin biosynthesis as a whole. Substantial differences, where they arise, will also be discussed. Despite these similarities, these pathways give rise to a diverse family of natural products as highlighted in Fig. 1.2[21,27], a feature which has attracted considerable attention.

## 1.3 Patellamides

Patellamides are cyclic octapeptides produced by *Prochloron didemni*, cyanobacteria that live symbiotically within the ascidian *Lissoclinum patella*[28]. Distinctive structural features include D-stereocentres, and thiazole and oxazoline heterocycles[27,28] (Fig. 1.3).

It is predicted that patellamides achieve a reduction in multidrug resistance, as described previously (section 1.2) through competitive binding to, and thus inhibition of P-glycoprotein (Pgp)[25]. Pgp is a transmembrane, ATP-dependent transport protein, frequently overexpressed in tumour cells[29], and is responsible for the efflux of a number of drug molecules, which ultimately leads to resistance[30]. Binding Pgp by patellamides therefore restricts the exclusion of active pharmaceuticals and so increases their efficacy in these resistant cell lines. The hypothesis that patellamide D targets Pgp was initially supported by the observation that cyclosporin A (Fig. 1.1), another cyclic peptide also binds Pgp[31]. More recently crystal structures of Pgp in complex with two cyclic peptides: cyclic-*tris-(R)*-valineselenazole (QZ59-RRR) and cyclic-*tris-(S)*-valineselenazole (QZ59-SSS) which are structurally related to patellamides (and cyanobactins more generally) have been solved[32] (Fig. 1.4). This reinforces the idea of an interaction between patellamide D and Pgp.

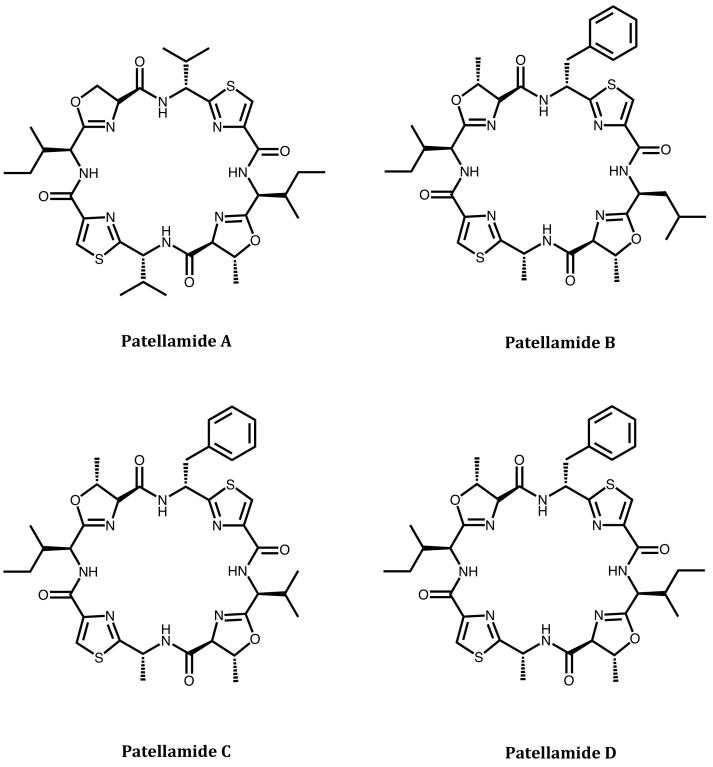



FIGURE 1.3: **Patellamides** Patellamides A, B, C and D. Trademark structural features include D-stereocentres, and thiazole and oxazoline heterocycles[27,28]

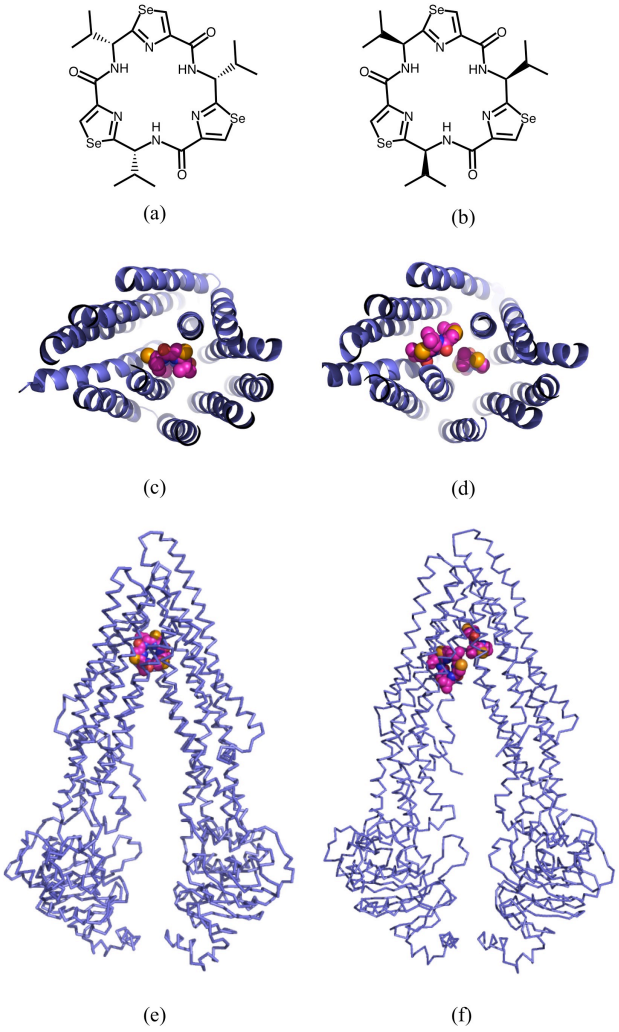



FIGURE 1.4: **Pgp in complex with cyclic peptides QZ59-RRR and QZ59-SSS** (a) structure of cyclic-*tris-(R)*-valineselenazole (QZ59-RRR). (b) structure of cyclic-*tris-(S)*-valineselenazole (QZ59-SSS). (c) QZ59-RRR and (d) QZ59-SSS bound at the top of the internal chamber of Pgp (blue cartoon, PDB: 4M2S, 4M2T). Alternate view of (e) QZ59-RRR and (f) QZ59-SSS. QZ59 compounds are shown as spheres coloured by element, where carbon is magenta, nitrogen is blue, oxygen is red and selenium is orange. N.B. The QZ59 compounds bind Pgp in slightly different locations and two molecules of QZ59-SSS are bound, compared with one molecule of QZ59-RRR. Pgp is shown as a blue ribbon.

## 1.4 Patellamide Biosynthesis

Patellamides, like other members of the cyanobactin and RiPP family of natural products, are synthesised ribosomally as a precursor peptide, which undergoes a number of enzymatic (and possibly non-enzymatic) chemical reactions during maturation to the

final macrocyclic product[27,28]. Their biosynthesis can be attributed to a single operon containing seven genes (*patA-G*) encoding proteins PatA through to PatG, including the precursor peptide (PatE) as well as the tailoring enzymes[28]. The PatE amino acid sequence can be seen in Fig. 1.5.

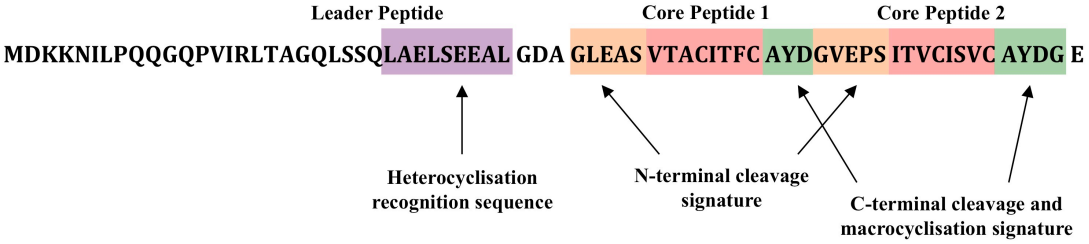



FIGURE 1.5: **Patellamide precursor peptide PatE** The gene product of *patE*, the precursor peptide (PatE) from which patellamides are derived is shown. The first 37 amino acids form the N-terminal leader peptide. Part of the leader peptide forms the recognition sequence for the heterocyclase (purple). The N-terminal cleavage recognition sequence (orange) and the C-terminal cleavage and macrocyclisation signature (green) flank the core peptide sequences (red).

The first thirty-seven residues of PatE constitute the leader peptide, a common feature of all RiPPs[20] (with bottromycins the only known exception, which contain a C-terminal follower in place of the N-terminal leader peptide[33,34]). The leader peptide increases the solubility of the precursor peptide and plays an important role in recognition of the PTM enzymes, and so is thought to be a requirement for processing[22,35]. Additionally it has been proposed that the leader peptide might act as a secretion signal, or perhaps provide a more protective and regulatory function, assisting in the folding of the precursor peptide, or inhibiting interaction with the PTM enzymes until biosynthesis of the secondary metabolite is required[22,35]. However at present, little evidence exists to support these proposals[22,35]. In the patellamide precursor, the leader peptide is followed by two, eight-residue core peptides, which go on to become two separate petellamides. Each core peptide is situated between an N-terminal protease sequence and a C-terminal protease/macrocyclisation signature[36]. The core peptide sequence itself is hypervariable[37].

Of the remaining genes in the operon *patA, D, E, F* and *G* are essential for *in vivo* patellamide synthesis[28,38,39]. The genes *patB* and *patC* have been reported as non-essential to synthesis and are at present of unknown function[28,38,39]. The gene products of the essential *patA, D, F* and *G*, and their roles in patellamide maturation will be

discussed in detail subsequently. A schematic summarising patellamide biosynthesis can be seen in Fig. 1.6. Cyclodehydration of core peptide cysteine, and serine/threonine residues generating thiazoline and oxazoline five-membered rings by PatD has been identified as the first step[27,40,41]. Once complete, the heterocyclised precursor peptide is cleaved at the N-terminus of the core peptide by PatA, removing the leader peptide[36,42]. The short, modified-precursor peptide is then simultaneously cleaved at the C-terminus of the core peptide and macrocyclised by PatG[36,43,44]. PatG also oxidises thiazolines to thiazoles[44]. The inversion of the stereochemistry of the two amino acids N-terminal to the thiazoles has not been attributed to a gene product of the biosynthetic operon and is generally thought to be spontaneous[45]. PatF has been identified as a prenyl-transferase. Although PatF is reportedly essential for the production of patellamides *in vivo*[46] it is not a requirement for their synthesis *in vitro*[47].

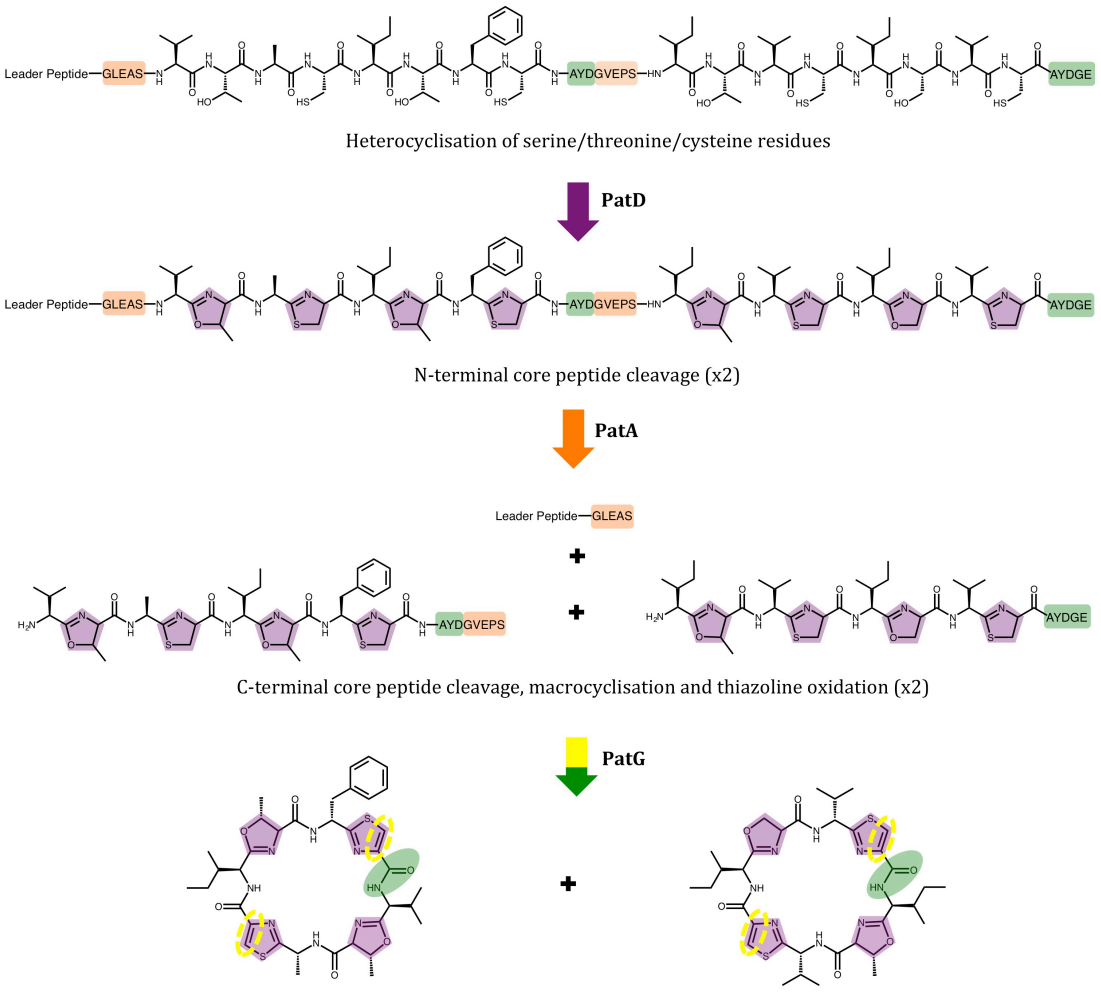


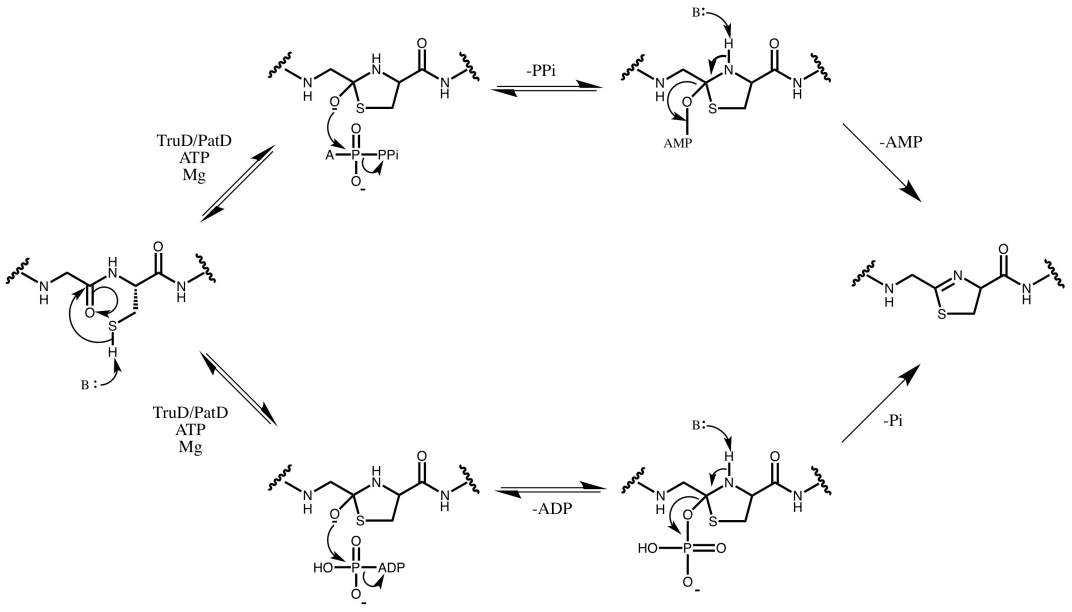

FIGURE 1.6: **Schematic of patellamide biosynthesis** Heterocyclisation of cysteine and serine/threonine residues to thiazolines and oxazolines respectively (purple) by PatD has been identified as the first step. Once complete, the heterocyclised precursor peptide is cleaved at the N-terminus of the core peptide, after the conserved recognition motif (orange), removing the leader peptide. The resulting precursor peptide fragments are macrocyclised by PatG, whilst simultaneously cleaving the core peptide from the C-terminal conserved recognition signature (green). PatG also oxidised thiazolines to thiazoles (yellow). It is not yet clear when during patellamide maturation oxidation occurs, only that it must follow heterocyclisation. In addition, the amino acid residues directly N-terminal to the thiazole heterocycles are epimerised. From our current understanding it is unclear whether this is spontaneous or enzymatic, or when during biosynthesis it occurs.

### 1.4.1 PatD: Heterocyclisation

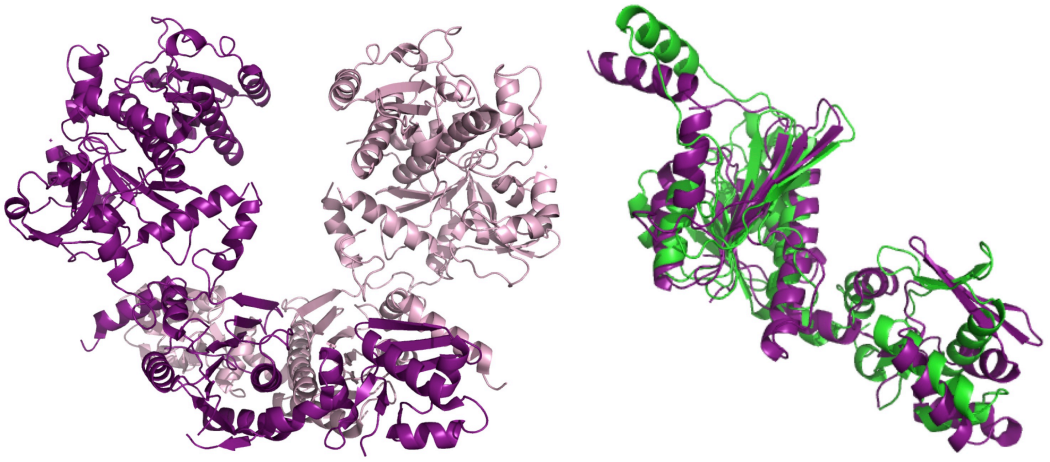
PatD is an ATP-dependent heterocyclase (cyclodehydratase), catalysing the formation of oxazoline and thiazoline rings, from serine/threonine and cysteine residues respectively[40,41,48]. In contrast, products of the trunkamide pathway contain only thiazoline heterocycles[21,39], despite the trunkamide heterocyclase (TruD) sharing high sequence homology (93 %) with PatD[40]. The reason for this difference in reactivity is not currently understood. The exact mechanism of heterocyclisation, particularly regarding the role of ATP during turnover remains disputed. An initial hypothesis proposed by McIntosh *et al.* was that both TruD and PatD function as molecular machines[48]. Later work, from Dunbar *et al.* on the homologous heterocyclase complex, BalhCD, from linear azole-containing peptide (LAP) biosynthesis, provided evidence for heterocyclisation proceeding via a phosphorylation mechanism[49,50]. The structure of TruD determined to 2.9 Å by Koehnke *et al.* revealed similarity between TruD and MccB, an adenylase from the microcin C7 antibiotic biosynthesis pathway (Fig. 1.7). This observation coupled with the observation that pyrophosphate is produced during turnover, lead Koehnke *et al.* to propose that TruD (and thus PatD) operate by an adenylation mechanism[51]. Therefore it is clear that further work is required to resolve these mechanistic ambiguities. Plausible phosphorylation and adenylation mechanisms can be seen in Fig. 1.7.

An attractive aspect of the cyanobactin heterocyclases from a biotechnological standpoint is their ability to process cysteine (and in some cases serine and threonine) seemingly irrespective of the immediately flanking residues[19,37,51]. The natural hypervariability of the cyanobactin core peptide sequences points to an interaction between the heterocyclase enzyme and some of the highly conserved regions of the precursor peptide, most likely within the leader peptide. If the leader peptide is removed from the precursor peptide, prior to treatment with PatD/TruD, heterocyclisation was found to be slow and inefficient[51]. A series of truncations and point mutations of the leader peptide of both PatE and TruE helped to identify the minimal portion of the leader peptide thought to be required for efficient processing by the heterocyclases[41,51,52]. This region, denoted 'minimal leader', spanning residues L26 through to L34 is therefore suspected to control recognition between the enzyme and the substrate precursor peptide[51,52] (shown in purple in Figs. 1.2, 1.5). In this way recognition is spatially distinct from catalysis,

explaining the observed tolerance of cyanobactin heterocyclases towards a variety of core peptide sequences. A molecular rational behind this substrate recognition was not established.




(a)

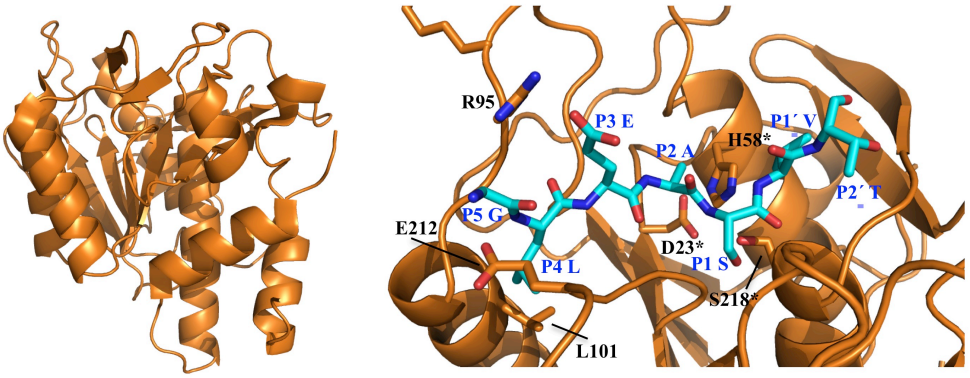


(b) (c)

FIGURE 1.7: **Heterocyclase mechanism and structure** (a) Two possible mechanism for heterocyclisation have been proposed. Both mechanisms proceed via the reversible formation of a hemiorthoamide intermediate, which ultimately eliminates to give a heterocycle. The proposed mechanisms differ regarding the utilisation of the nucleotide in the activation of the hemiorthoamide intermediate with evidence supporting an adenylation mechanism (top) and a kinase mechanism (bottom). (b) X-ray crystal structure of TruD dimer (PDB: 4BS9) represented as a cartoon (purple for one monomer and pink for the other). (c) structural comparison between domains 1 and 2 of TruD (purple) and MccB (green, PDB: 3H5A) and adenylase from C7 antibiotic biosynthesis.

### 1.4.2 PatA: N-terminal Core Peptide Cleavage

PatA is 75 kDa protein consisting of an N-terminal subtilisin-like serine protease domain (PatApr) and a C-terminal domain of unknown function (PatA-DUF)[36]. The PatApr domain catalyses the removal of the leader peptide from PatE, via proteolytic cleavage between the recognition sequence, N-terminal to each core peptide, and the first residue of the core peptide[36]. While the recognition sequence G(L/V)(D/E)(A/P)(S/P) is highly conserved and necessary for PatApr activity, there appears to be no conservation of the first residue of the core peptide[42]. The crystal structure of the protease domain has been solved, independently by two labs[42,44] (Fig. 1.8). Modelling of the peptide GLEASVT (P5-P2′) into the active site of PatApr rationalised the conservation of the N-terminal recognition signature throughout cyanobactin bionsynthesis, and the activity of the enzyme regardless of the succeeding core peptide sequence[42] (Fig. 1.8). Hydrolysis of the amide bond proceeds via nucleophilic attack by S218, part of an D-S-H catalytic triad, typically found in subtilisin-like serine proteases[53]. Proteolysis *in vitro* was reported to be very slow, with a turnover of <10 substrate molecules a day[36,42]. It was suggested a slow turnover might be advantageous *in vivo*, allowing for complete processing by the heterocyclase before the leader peptide is removed[48]. A recent study has shown the relative rates of the proteolysis and heterocyclisation reactions can be controlled by altering the redox potential of the reaction[54]. The precursor peptide must be reduced to react with the heterocyclase (free cysteine is the substrate of the heterocyclase), whereas non reducing conditions correlate with an increase in protease activity[54]. If both PatA and ThcD (TruD homologue) are added simultaneously to the precursor peptide in the presence of reducing agent, then the desired heterocyclised, cleaved intermediate is produced as the major product. Conversely, in the absence of reducing agent, the non-heterocyclised, cleaved intermediate is observed as the major product, consistent with PatA acting before ThcD[54]. It has been proposed that it is by controlling the redox potential, and thus the relative rates of the heterocyclase and the N-terminal protease that cyanobacteria establish control of product formation *in vivo*[54]. It is important to note that the PatApr domain is active *in vitro* in isolation, and it doesn't require the presence of the C-terminal DUF. It is currently unclear, what effect, if any, the DUF has on PatApr activity. The DUF will be discussed in more detail separately.




(a) (b)

FIGURE 1.8: **PatApr** (a) X-ray crystal structure of PatApr (PDB: 3ZXX) shown as an orange cartoon. (b) PatApr (orange) with modelled substrate (cyan sticks). PatApr residues involved in substrate recognition are shown as orange sticks. Catalytic residues are marked with an asterisk.

### 1.4.3 PatG: Oxidation and Macrocyclisation

PatG is a 131 kDa multi-domain protein. Metagenomic analysis identified a flavin mononucleotide (FMN) binding site towards the N-terminus of PatG, indicative of oxidase activity (PatGox)[44]. More often than not in cyanobactin biosynthesis, like in the patellamide example, the oxidase domain is located at the N-terminus of the G protein. However, there are examples, such as the cyanotheceamide biosynthetic pathway, where the oxidase is found as a freestanding protein[26]. Oxidase activity was confirmed when the cyanothecamide oxidase was shown to catalyse thiazoline oxidation on a macrocyclic substrate, while the oxidase from the arthrospiramide pathway affected both linear and macrocyclic substrates *in vitro*[47]. However, it has not been confirmed in nature. Interestingly, while other cyanobactins, including the tenuecyclamides and microcyclamides, contain both oxazoles, and thiazoles, only patellamides containing oxazolines and thiazoles have been identified[16,21,26]. It seems likely therefore, that similarly to the heterocyclases, different oxidase domains from various cyanobactin biosynthetic pathways, despite their high sequence homology are capable of different chemistries. For example, through analysis of the products of aestuaramide (*lyn*), cyanotheceamide (*cya*), arthrospiramide (*art*) and microcyclamide (*mic*) pathways (Fig. 1.2), it is possible that MicGox catalyses the oxidation of both thiazolines and oxazolines, while Lyn, Cya and ArtGox (in addition to PatGox) only oxidase thiazolines (Fig. 1.9).

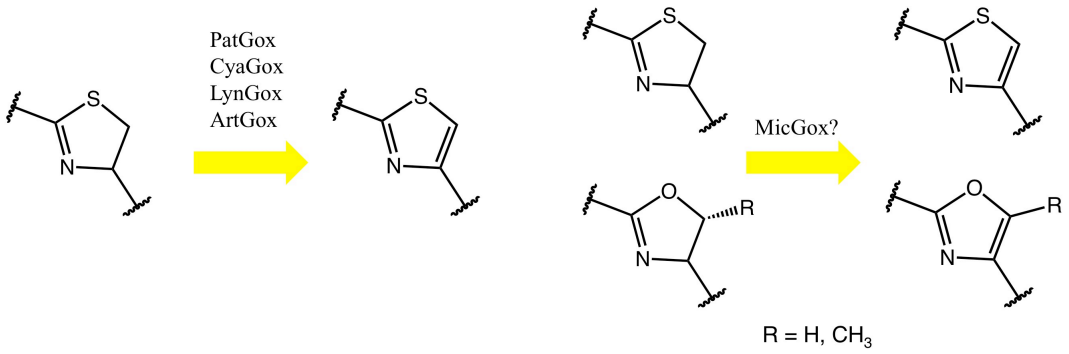
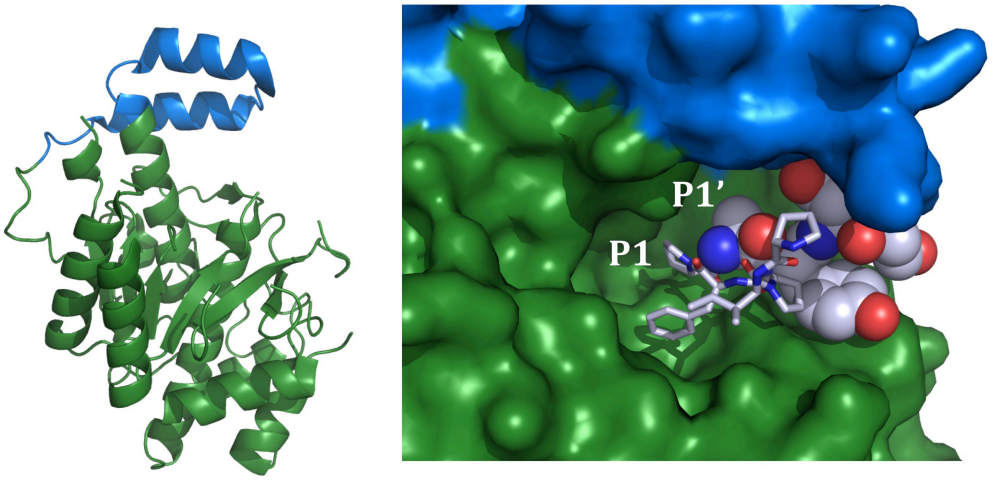



FIGURE 1.9: **Cyanobactin oxidation** The N-terminal domain of the G enzymes (Gox) is responsible for heterocycle oxidation. Reflecting the observed difference in activity of the D enzymes of different pathways, it is expected, through analysis of the products of the patellamide, aestuaramide, cyanothecamide and microcyclamide pathways that Pat, Lyn Cya and ArtGox catalyse thiazoline oxidation, while MicGox catalyses both thiazoline and oxazoline oxidation.

In addition to an oxidase domain, PatG also contains a subtilisin-like serine protease domain (PatGmac). Despite the similarity between the protease domains of PatA and PatG[36], the enzymes act on different substrates, with PatG cleaving the core peptide from the C-terminal (A/S)(Y/F)DG recognition motif, subsequent to the removal of the leader peptide[43,55]. Furthermore, the PatG protease domain has been found to be solely responsible for the macrocyclisation reaction[43,44,55].

The crystal structure of PatGmac in complex with substrate peptide VPAPIPFPAYDG (P8-P4′) was solved, providing valuable insight into substrate recognition[43] (Fig. 1.10). Recognition is achieved entirely through interactions between the enzyme and the conserved 'AYDG' motif, and not with the core peptide[43]. The outcome is a promiscuous enzyme that will accept a wide variety of substrates, with varying compositions of the core peptide, including unnatural amino acids, providing the C-terminal recognition element is present[55–58]. However there is one limitation: a heterocycle (or a proline in undmodified substrates) is required at the final position (P1) of the core peptide to angle the peptide away from steric clashes with the enzyme[43].

The PatGmac-peptide complex structure can also be used to hypothesise the mechanism of macrocyclisation (Fig. 1.11). The reaction proceeds via attack of S783 on the C-terminal carbonyl of the core peptide, creating an acyl-enzyme intermediate. The acyl-enzyme intermediate has been observed by mass spectrometry[43]. At this point, the




(a) (b)

FIGURE 1.10: **PatGmac structure** (a) X-ray crystal structure of PatGmac represented as a cartoon (green) with the macrocyclisation insert highlighted in blue. (b) Close-up view of the active site of PatGmac (shown as surface) in complex with VPAPIPFPAYDG substrate mimic. The N-terminal residues (VPA) are disordered in the structure. The remaining core peptide residues (PIPFP) are shown as white sticks, while the recognition element (AYDG) is shown as white spheres. This representation shows that where the acyl-enzyme is formed is shielded from water by the macrocyclisation insert (blue surface) and the AYDG peptide.

acyl-enzyme intermediate can be hydrolysed by either water, releasing linear peptide, or by the free N-terminus of the core peptide, releasing the desired macrocyclic product. The inherent *cis* conformation of the heterocycle at P1 orients the N-terminus of the peptide for nucleophilic attack of the acyl-enzyme intermediate, resulting in head-to-tail (N-C) cyclisation[43]. A large helix-loop-helix insertion (often referred to as the macrocyclisation insert) adjacent to the active site shields the acyl-enzyme intermediate from water, thus favouring macrocyclisation over hydrolysis[43]. It has also been hypothesised that if the AYDG motif were to remain bound in the active site-post cleavage, the steric bulk of these residues would assist in shielding the acyl-enzyme intermediate[43].

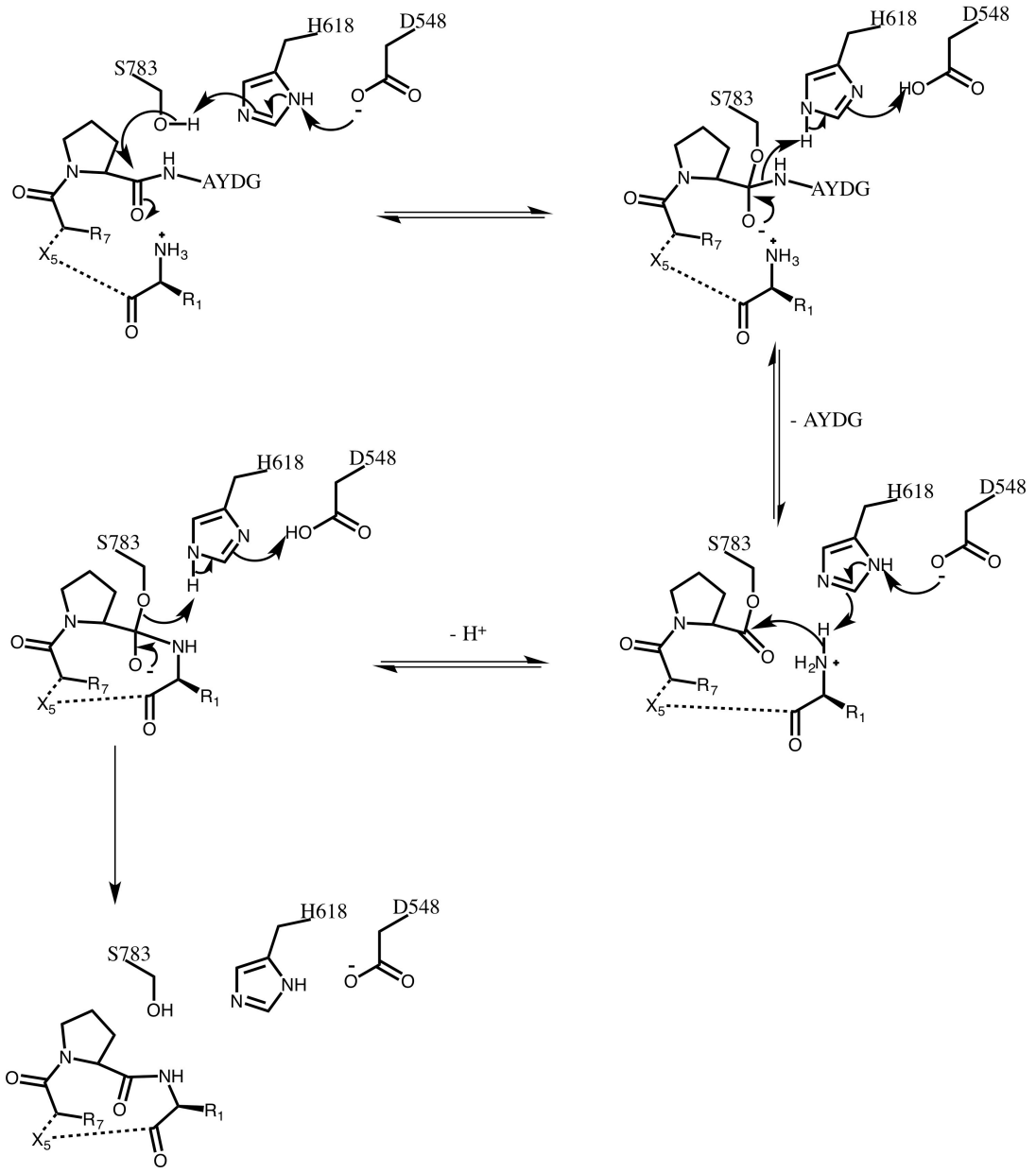



FIGURE 1.11: **Macrocyclisation mechanism** Proposed mechanism for macrocyclisation: The active site serine attacks the proline carbonyl adjacent to the AYDG motif creating a characteristic acyl-enzyme intermediate, releasing AYDG. The incoming N-terminus of the core peptide sequence displaces AYDG and hydrolyses the acyl-enzyme intermediate forming the macrocycle.

Finally, PatG harbours a C-terminal domain of unknown function (PatG-DUF) homologous to the DUF from PatA[44]. Like PatA, the protease domain of PatG is active *in vitro* in the absence of the DUF, and so the role of the DUFs in patellamide biosynthesis is to be investigated.

### 1.4.4 PatF: An Inactive Prenyl-transferase

Sequence homology revealed the PatF family of PTM enzymes to be prenyl-transferase enzymes[57,59,60]. Interestingly, while prenylation is a common modification found in cyanobactins, no prenylated patellamides have been discovered[27]. Activity has been demonstrated for the homologous cyanobactin prenyl-transferases TruF1 and LynF[57,59], but PatF appears to be inactive[60] (Fig. 1.12). TruF1 catalyses the forward-O-prenylation of serine and threonine residues, whereas LynF is responsible for the reverse-O-prenylation of tyrosine, which spontaneously undergoes a Claisen rearrangement to give a forward-C-prenylated product.[59] (Fig. 1.12). The structure of PatF revealed that mutations of two catalytically important residues explain the inactivity of PatF and thus the absence of prenylated natural products from the patellamide pathway[60] (Fig. 1.12). Attempts to mutate PatF to restore activity resulted in insoluble protein[60]. Whether PatF has another role in patellamide biosynthesis, and if not, why it is conserved in the patellamide gene cluster is unknown[27].

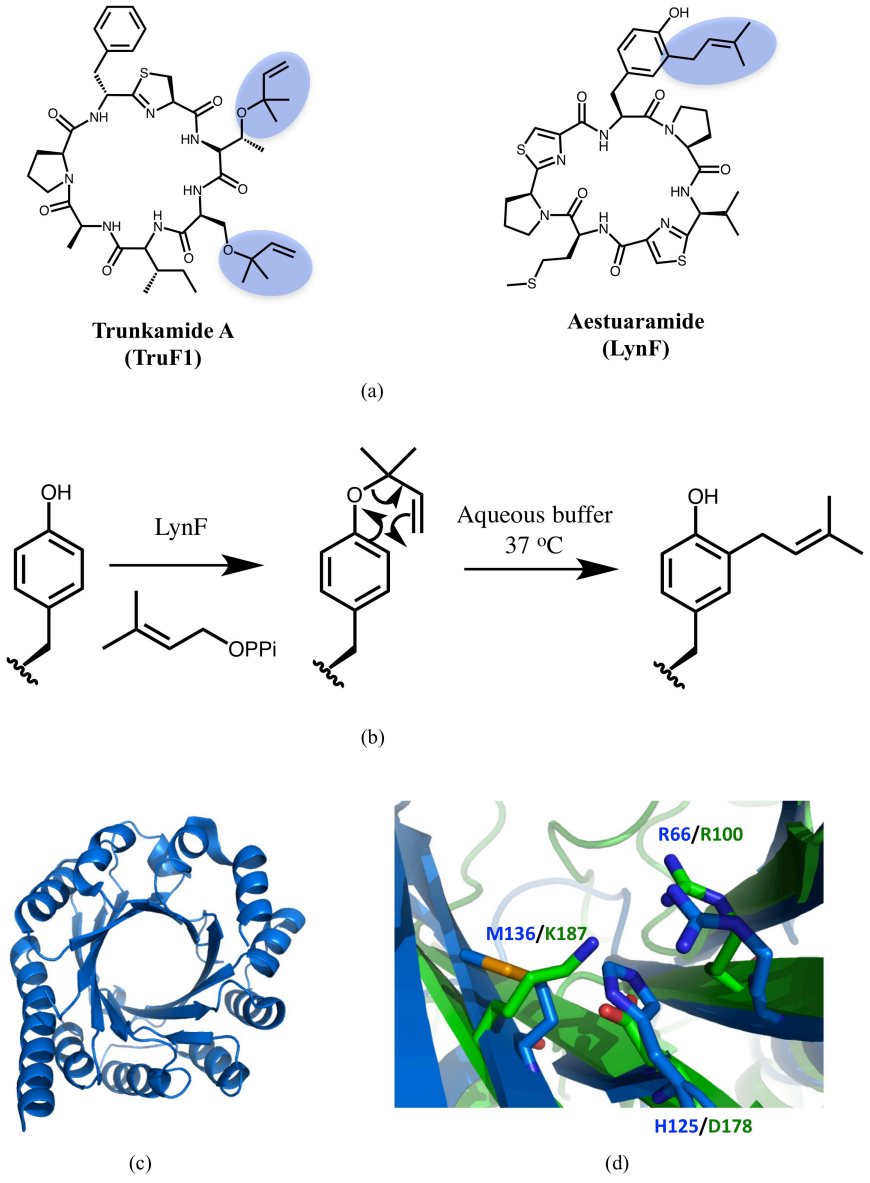



FIGURE 1.12: **PatF: inactive prenyl-transferase** (a) Example structures of a serine-O-prenylated cyanobactin trunkamide A (catalysed by TruF1) and a tyrosine-C-prenylated cyanobactin aestuaramide (catalysed by LynF). (b) LynF catalyses the reverse-O-prenylation of tyrosine. A spontaneous Claisen rearrangement gives the forward-C-prenylated product. (c) X-ray crystal structure of PatF (PDB: 4BG2) represented as a cartoon (blue). (d) Overlay of PatF (blue) with known prenyl-transferase DMATS (green). Important active site residues are shown as sticks. K136M and H125D mutations (PatF numbering) are thought to impair substrate binding, thus inactivating PatF [60].

### 1.4.5 Epimerisation

Extensive structural analysis of patellamide D and lissoclinamides 4, 5 and 6, confirmed that the amino acids directly N-terminal to the thiazole/thiazoline rings are epimerised to D-stereocentres[61]. Additionally many of the related cyanobactins that contain heterocycles have also been reported to contain D-amino acids, including trunkamide A[62,63], lissoclinamide 7[45,61] and the microcyclamides[64]. For the teneucyclamide and trichamide families of cyanobactins, the presence of D-amino acids has been proposed, but has not been confirmed in all products[65,66]. Since ribosomes are selective for L-amino acids, epimerisation must be a PTM[45].

From our current understanding of the biosynthetic pathway, we cannot state definitively when, during patellamide maturation, epimerisation takes place. Epimerisation can be in theory either acid or base catalysed. While the α-carbon adjacent to the thiazoline is not particularly acidic, perhaps favouring acid catalysis over base catalysis, thiazolines are chemically unstable under acidic conditions and readily ring open to give cysteine[50]. Assuming then a base catalysed mechanism, epimerisation would seem favourable after thiazoline formation, and prior to its oxidation to a thiazole, thus avoiding the need to disrupt aromaticity to stabilise the build up of negative charge that occurs following proton abstraction. Therefore the following base catalysed mechanism, where proton abstraction is stabilised through conjugation onto the imine nitrogen seems viable[45] (Fig. 1.13).

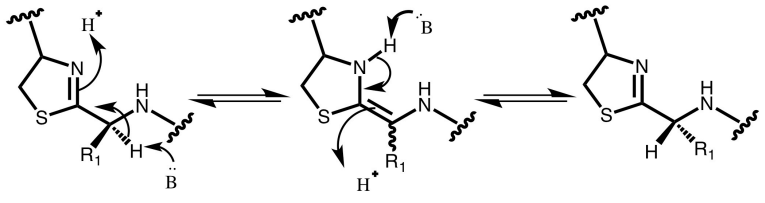



FIGURE 1.13: **Epimerisation mechanism** A base abstracts a proton at the α-carbon adjacent to the thiazoline ring. This is stabilsed via conjugation and subsequent protonation of the imine nitrogen. Re-protonation occurs at the α-carbon on the opposite face.

Since it is not explicit whether oxidation directly succeeds heterocyclisation, follows leader sequence removal, or only occurs on the cyclic peptide, the precise timing of

epimerisation is uncertain. Furthermore it is unclear as to whether epimerisation is a spontaneous or an enzymatic process in cyanobactin biosynthesis. Early work by Wipf *et al.* demonstrated that if lissoclinamide 7 was synthesised with L-valine adjacent to the thiazoline, it would epimerise to D-valine following treatment with pyridine[45] (Fig. 1.14), consistant with the $\alpha$-carbon adjacent to a thiazoline being susceptable to non-enzymatic epimerisation.

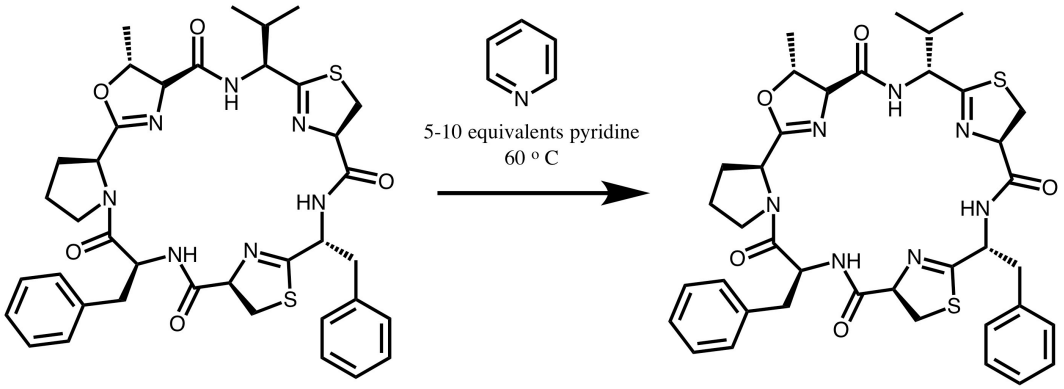



FIGURE 1.14: **Epimerisation of lissoclinamide 7** Epimerisation of unnatural L-isomer to the natural D-isomer using 5-10 equivalents of pyridine.

However, the required 5-10 equivalents of pyridine and temperatures of 60 °C to afford the correct isomer[45] do not reflect the physiological conditions of the cyanobacteria, and so it cannot be considered proof of spontaneous epimerisation in nature. More convincing was the observation by Salvatella *et al.* that an enantiomer of trunkamide A synthesied with L-phenylalanine, when solubilised in a combination of 30 % d6-DMSO/70 % $CDCl_3$ for a series of NMR experiments, epimerised to the natural D-enantiomer within days[67] (Fig. 1.15). Analyses of solution structures of the isomers revealed trunkamide A undergoes a significant conformational rearrangment following inversion of the F residue, resulting in a less sterically hindered, planar structure and allows for two, new hydrogen bonding interactions[67] (Fig. 1.15).

It was proposed that this increase in conformational stability of the macrocycle is the driving force for the epimerisation reaction[67]. This agrees with epimerisation being a spontaneous process, and that if it is, it most likely occurs on the macrocycle and not on the linear peptide. It is difficult to rationalise a similar thermodynamic driving force for a spontaneous epimerisation on the flexible linear peptide. Such a

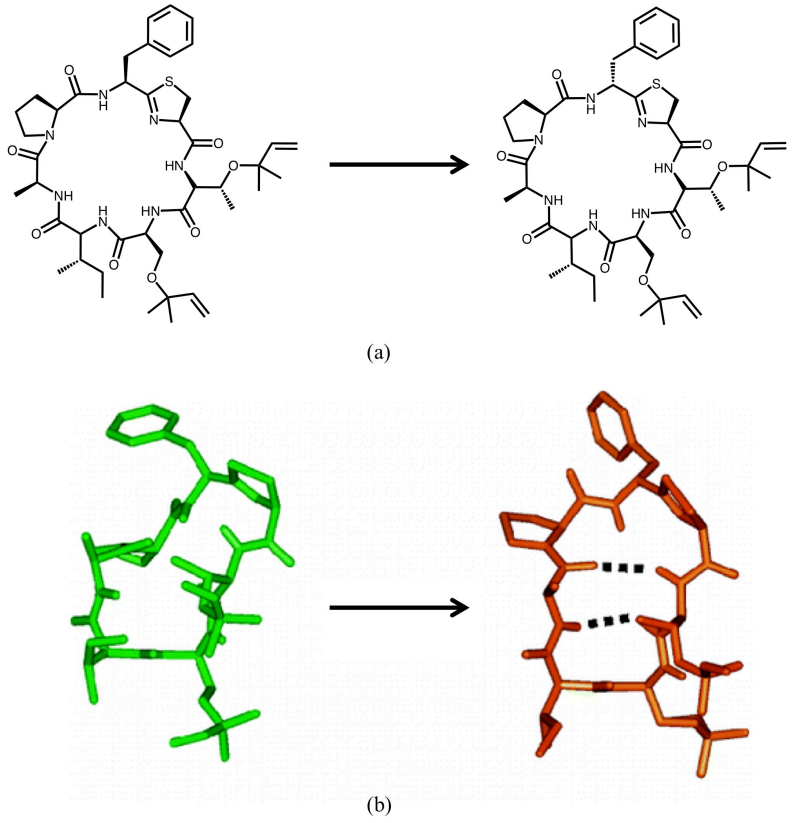



FIGURE 1.15: **Epimerisation of trunkamide A** Unnatural L-isomer spontaneously epimerises to natural D-isomer within days. (a) 2D stick representation; (b) 3D solution NMR structures determined by Salvatella *et al.* (figure taken from paper)[67]

'conformational stability' hypothesis might be sufficient to explain the conservation of D-stereocentres throughout the cyanobactin superfamily[21] and why the natural compounds are isolated as single enantiomers and not diastereoisomers[45]. That said, cyanobactins are structurally highly diverse, with known compounds of different ring sizes and different constitutent amino acids. Therefore, it might not be that every cyanobactin will experience the same conformational pressures favouring epimerisation, as observed for trunkamide A. D-Stereocentres have so far always been reported adjacent to thiazoline/thiazole heterocycles, and never next to oxazoline/oxazole heterocycles, even though these residues are also potentially labile to epimerisation. Such selectivity, whether chemical or regioselectivity, might be under control of an enzyme catalyst. While there is evidence supporting a spontaneous epimerisation of synthetic cyanobactins, the requirement of an epimerase in their biosynthesis has not been explicitely ruled out. However no epimerase activity has been identified in any of the biosynthetic enzymes, nor is epimerase activity predicted based on sequence homology of

the remaining uncharacterised gene products in the biosynthetic cluster[27]. Epimerase, racemase and isomerase enzymes are common place throughout biology, exhibiting a vast diversity in structure and size, and subtle variation in their catalytic mechanisms, especially concerning the amino acid combination utilised as acids and bases[68–72]. This makes it difficult to predict epimerase activity amongst the uncharacterised proteins in cyanobactin pathways. If epimerisation is catalysed by one of these uncharacterised gene products, this would potentially represent a novel class of epimerase.

### 1.4.6 Proteins of Unknown Fucntion

PatA and PatG both contain homologous (56 % identical) C-terminal DUFs. In patellamide biosynthesis, the presence of two DUFs and two epimerised amino acids might correlate, raising the possibility of the DUFs catalysing epimerisation. The conservation of the DUFs throughout cyanobactin biosynthetic pathways suggest they play a role in their biosynthesis, but they do not appear necessary for their synthesis *in vitro*[47]. Their role in biosynthesis is still to be determined.

In addition to the C-terminal domains of PatA and PatG, there exist two hypothetical proteins of unknown function, PatB and PatC. Like the DUFs, these proteins are associated to and conserved throughout cyanobactin biosynthetic pathways[26]. Neither protein is necessary for their synthesis *in vitro*[47] or *in vivo*[28] and the reason for their conservation is also unknown.

## 1.5 Exploiting the Advantageous Properties of Cyanobactin Biosynthesis for the Development of Cyanobactin Analogues for Biological Application

Despite the efficacious, clinically-relevant bioactivities of cyanobactins, coupled with the pharmacological advantages of cyclic peptides as therapeutics (as discussed at the beginning of this thesis (Section 1.1), the development of these compounds as pharmaceutical lead compounds has been slow, largely due to a low availability of material from their natural sources. Creating a large variety of macrocycles in sufficient

yields is paramount in ascertaining structural activity relationships of these bioactive molecules, and for their subsequent development as drugs.

### 1.5.1 Chemical Synthesis

The typical approach is to turn to synthetic chemistry to access greater amounts of material for subsequent testing. However, as with many natural products, their complexity makes them challenging targets for *de novo* synthesis[17]. One of the most efficient syntheses of cyanobactins was that of patellamide A by Garcia-Reynaga *et al.* in 2008, forming patellamide A from a 14 step reaction with an overall yield of 55 %[73]. While this synthesis boasts modest yields and relatively mild reaction conditions, it utilises a patellamide A-specific starting point - a valine-thiazole dipeptide for which the synthesis was previously established[74]. Therefore there is no guarantee the synthetic strategy will be as efficient in making the other patellamides, or that it will be applicable to the synthesis of diverse range of cyanobactins and cyanobactin-like compounds. Although the diversity offered through synthetic chemistry is generally considered to be unlimited, the potential requirement to alter the synthetic strategy for each macrocycle would prove expensive and time consuming.

### 1.5.2 *in vivo* Synthesis

A diverse range of cyanobactins can be produced *in vivo*, utilising cultured bacteria as 'chemical factories'. While so far attempts to culture *Prochloron* sp. have been unsuccessful[75] products of the *tru* pathway have been produced recombinantly in *Escherichia. coli* (*E. coli*)[38]. The promiscuity of the cyanobactin biosynthetic enzymes mean that simply by engineering the TruE precursor peptide a huge variety of compounds can be produced in a straightforward and cost effective manner. An extensive study by Ruffner *et al.* utilised a NNK (where N = A/C/G/T and K = G/T)[76] mutagenesis strategy to modify the *truE* gene to assess the combinatorial potential of the *tru* pathway[56]. In this library design, to account for the requirement of a C-terminal heterocycle to facilitate macrocyclisation, cysteine was fixed at position 7 (N to C-terminal notation). In addition, proline was fixed at position 5, as it was reasoned to be important for the structure of the final compound[56,67]. Even with proline and cysteine fixed, a full degenerate XXXXPXC library contains 3.2 million

possible sequences. In order to determine allowed sequences, that are processed by the *tru* pathway, three double mutant (XXIAPFC, TSXXPFC and TSIXPXC) and a quadruple mutant library (XSXXPXC) were designed, so that all positions other than 7 and 5 were randomised in at least a subset of libraries. These smaller libraries simplified analysis, but still demonstrated the huge combinatorial potential of the *tru* pathway, with >300 unique, new compounds[56]. Furthermore the modifying enzymes have been shown to tolerate unnatural amino acids, expanding the accessible diversity[57]. The non-proteinogenic amino acids are introduced using an orthogonal tRNA/aminoacyly-tRNA synthetase pair that adds a specific unnatural amino acid in response to a nonsense frameshift mutation[57]; however, currently only a limited number of unnatural amino acids can be introduced incorporated, and only a single type of unnatural amino acid can be incorporated into a single polypeptide chain[77]. In general, major drawbacks of these methods are the low yields, with the macrocyclic products only being obtained in μg $L^{-1}$ quantities[57], and the difficulty in their subsequent extraction from large volumes of culture. However, a recent breakthrough in culture conditions, essentially the addition of $H_2S$, increased the cultured yield of patellins produced by the *tru* pathway, over 150 fold, and the addition of mevalonate increased the yield a further 18 fold[54]. The effect of these compounds was not, as first expected, due to an increase in the level of expression of the substrate peptide, or PTM enzymes. Instead, the authors concluded that sulfide regulated the activity of the enzymes, ensuring each one acted at the correct point in the sequence of biosynthetic events[54]. Mevalonate is a known precursor to dimethylallylpyrophosphate (DMAPP)[78], a substrate of the prenyl-transferase enzymes[57,59,60], and so it was assumed the increase in the yield upon mevalonate additon was due to an increase in activity of the prenyl-trasnferase. However, when the *pat* pathway, which contains an inactive prenyl-transferase homologue, was expressed, an increase in the cultured yield of patellins upon the addition of mevalonate was still observed[54]. Therefore, the reason for the increase in yield, following supplementation with mevalonate is unknown. If this approach is general, then it will transform *in vivo* production.

### 1.5.3 *in vitro* Synthesis

Cyanobactin biosynthetic pathways can be reconstituted *in vitro* enabling the production of natural and unnatural cyanobactins on a mg $L^{-1}$ scale[47]. By using isolated substrate

peptides and enzymes from multiple cyanobactin biosynthetic pathways, the *in vitro* synthesis is robust and scalable. As with the *in vivo* synthesis, the use of biosynthetic enzymes ensures the process is green and keeps synthetic costs to a minimum. The *in vitro* synthesis makes use of engineered precursor peptides with varying core peptide sequences. The modified precursor contains a single core peptide sequence and a C-terminal $His_6$ tag to aid purification. A lysine residue inserted between the N-terminal recognition sequence 'GLEAS' and the start of the core peptide sequence facilitates the use of trypsin as a more rapid alternative to PatA cleavage (Fig. 1.16). Such peptides can be overexpressed in *E. coli* at high yields (50-200 mg $L^{-1}$) [47]. The enzymes are also accessible via straight forward expression and purification at high yields [47]. Diversity can be created, not only through the use of various substrate peptides, but by combining substrates with different combinations of enzymes from different pathways [47]. In this way multiple products can be produced from a single precursor peptide (Fig. 1.16). Such an engineered precursor peptide with the core peptide sequence ITACITFC (coding for the natural product patellamide D) is usually employed as a standard substrate throughout this thesis and is simply referred to as PatE′.

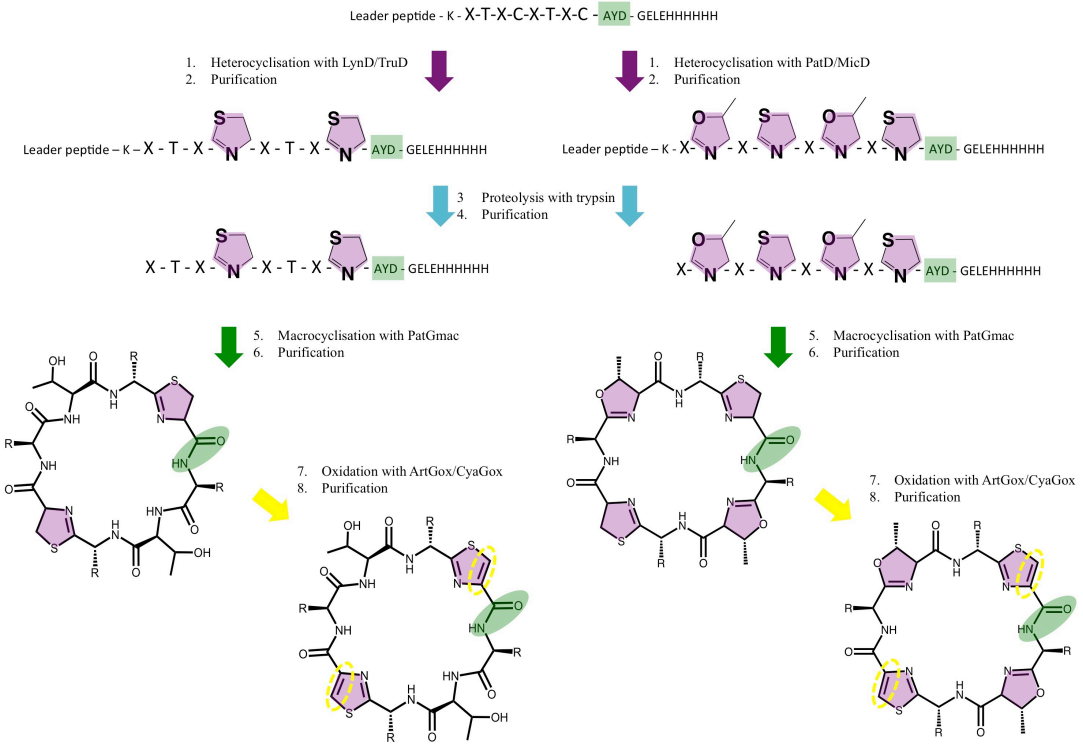



FIGURE 1.16: **A schematic representation of the *in vitro* synthesis of azol(in)e containing cyclic peptides** A full description of the *in vitro* synthesis can be found elsewhere (Chapter 5). Briefly: Engineered PatE′s with differing core peptide sequences (where X can be any amino acid (except lysine or arginine due to the use of trypsin to remove the leader sequence)) can be processed with either a cysteine heterocyclase (TruD/LynD) or a cysteine, threonine and serine heterocyclase (PatD/MicD). Only the C-terminal cysteine needs to be conserved for subsequent processing by PatGmac - all the other positions are in principle variable. The heterocyclised PatE′ is then purified and incubated with trypsin to remove the leader sequence. The cleaved, heterocyclised PatE′ is purified away from trypsin and macrocyclised by PatGmac. The purified macrocycle can then be reacted with either ArtGox or CyaGox to to oxidise thiazolines to thiazoles. In this way, using just the enzymatic steps described a minimum of four macrocyclic products can be produced from a single precursor peptide starting material.

The ability to produce a wide variety of highly modified cyclic peptides in good yields, as offered by the *in vitro* synthesis method, provides a more facile route to further testing and development, and might spark renewed interest and investigation into cyanobactins as potential therapeutic compounds.

## 1.6 Aims and Objectives

Manipulation of cyaonobactin biosynthetic enzymes for the production of a diverse range of azol(in)e containing macrocycles on a useful scale, as described for the *in*

*vivo* and *in vitro* syntheses, requires a comprehensive understanding of the individual enzymes involved. Much of cyanobactin biosynthesis has been characterised, however some unknowns and uncertainties remain. These studies, through a combination of structural biology, biochemical and biophysical characterisation, aim to address the current unknowns in patellamide biosynthesis, with a focus on using any new insight gained, coupled with our existing knowledge, to develop a biotechnological toolkit for the synthesis of cyanobactins and ultimately diverse cyclic peptides more generally, thus improving upon the existing *in vitro* synthesis.

There are three specific aspects of the biosynthesis that are investigated in this thesis: (1) Investigation into the role of the C-terminal DUFs - What is their role in nature? Are they involved in synthesis, either directly or indirectly? Do they constitute an essential, or beneficial part of the biotechnological toolkit? (2) Detailed characterisation of the heterocyclase enzyme - What is the mechanism of heterocyclisation? What is the molecular rational governing substrate recognition of the heterocyclase? (3) Investigation into epimerisation - When during synthesis does epimerisation occur? Is it spontaneous or enzymatic? Can the stereochemistry of these residues be controlled? Does an epimerisation step need to be built into the *in vitro* synthesis?

Finally, these studies aim to develop the *in vitro* synthesis further, in an attempt to improve the efficiency and flexibility of the process. Ultimately the *in vitro* synthesis will be tested for its applicability in producing a range of patellamide D analogues in an attempt to determine structural activity relationships for Pgp inhibition.

# Chapter 2

# Structural and Biochemical Characterisation of PatG-DUF

## 2.1 Introduction

The roles of the C-terminal domains of PatA and PatG in patellamide biosynthesis are unknown and hence they are designated as Domains of Unknown Function (DUFs). The high sequence homology between the domains implies their function is related (Fig. 2.1).

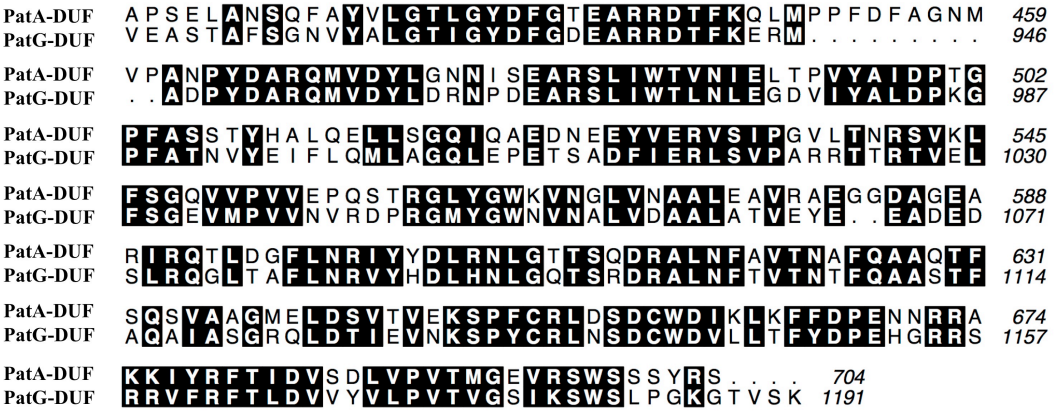



FIGURE 2.1: **Sequence alignment of PatA-DUF and PatG-DUF** The conserved C-terminal DUFs of PatA and PatG share 56 % identity. Fully conserved residues are shaded black and non-conserved residues are unshaded.

Moreover both DUFs are highly conserved in all cyanobactin pathways, (some of which can be seen in Figs. 2.2, 2.3), suggesting they have an important role in cyanobactin biosynthesis.

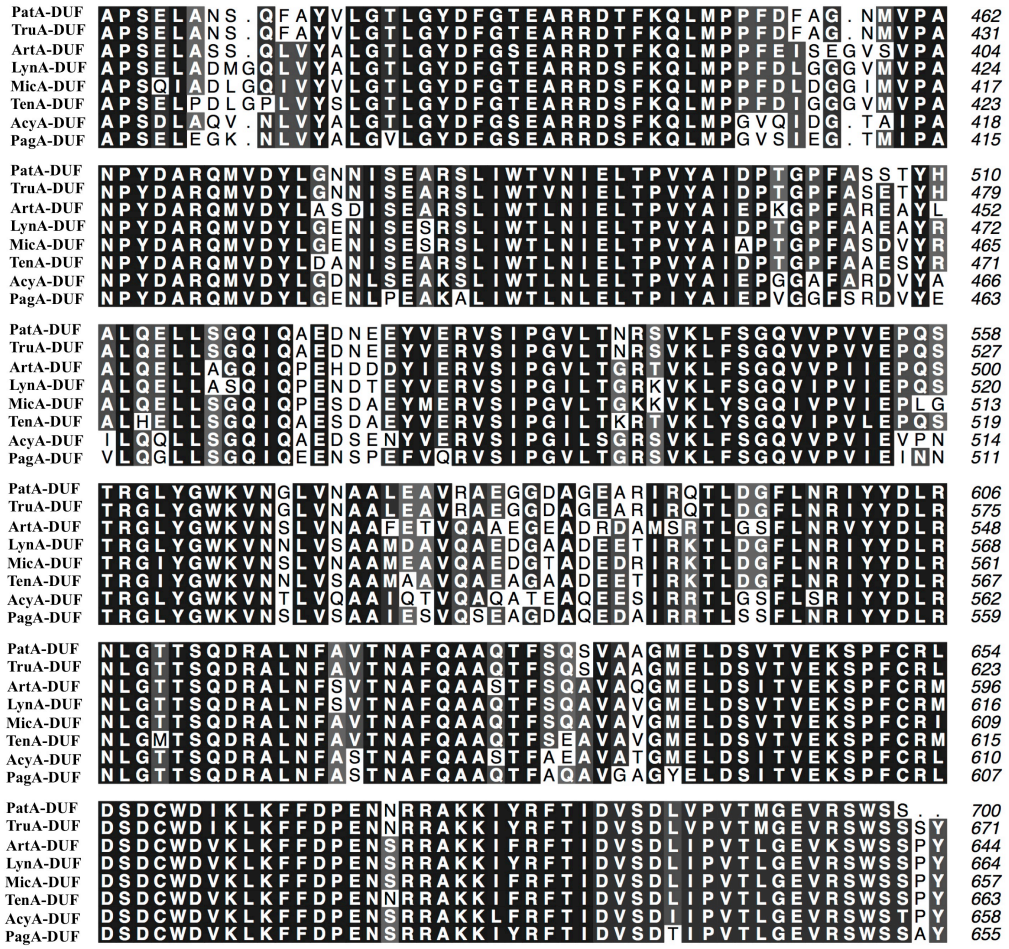



FIGURE 2.2: **PatA-DUF homologues** Sequence alignment between PatA-DUF and homologues from related cyanobactin pathways. Fully conserved residues are shaded black, partially conserved residues are shaded in grey, non-conserved residues are unshaded.

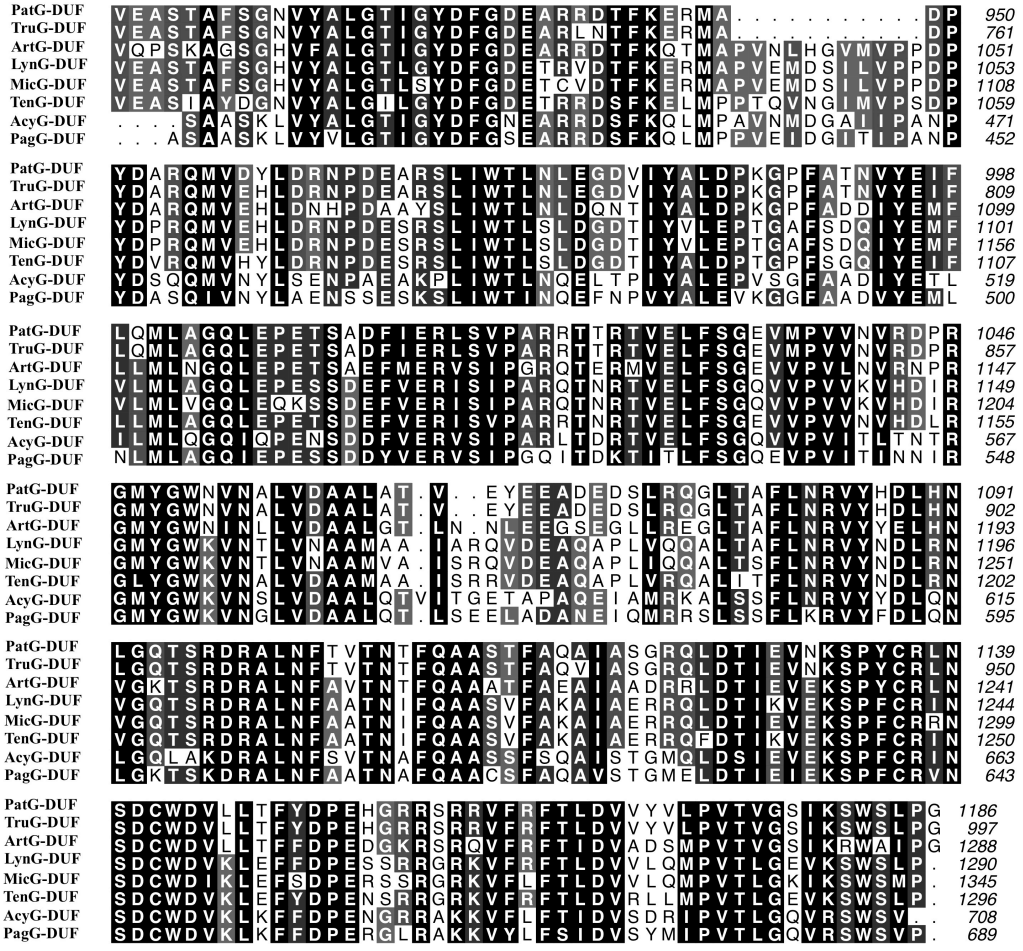



FIGURE 2.3: **PatG-DUF homologues** Sequence alignment between PatG-DUF and homologues from related cyanobactin pathways. Fully conserved residues are shaded black, partially conserved residues are shaded in grey, non-conserved residues are unshaded.

Broadly speaking, three possibilities for the DUFs exist: (1) In all examples, the C-terminal DUFs are associated with domains exhibiting protease activity. Therefore it is possible the DUFs affect proteolysis - either to accelerate reaction rates (PatApr turnover is very low in isolation[42]), or to regulate activity. (2) PatG is a multi-domain protein, containing an oxidase domain and a macrocyclase domain in addition to the C-terminal domain of unknown function. The oxidase and macrocyclase domains catalyse two, very different chemical reactions, and both domains are active in isolation[47]. Therefore it is plausible that the DUFs exhibit intrinsic catalysis, chemically distinct from that of their associated domains. In this case, epimerisation of the two residues N-terminal to the thiazole heterocycles is the most likely candidate, as it is the only post-translational modification that hasn't currently been attributed to a gene product of the patellamide

operon. (3) The DUFs might affect patellamide biosynthesis indirectly. They could provide a chaperone-like function, assisting in the pre-organisation of the precursor peptide prior to biochemical modification. Alternatively they might be involved in host-cell immunity, and/or export of the cyclic peptides from the host-cell. It is also possible they are involved in mediating PPIs, providing a scaffold for multi-enzyme complex formation.

Since one of the on-going goals of the lab is to optimise the established *in vitro* patellamide synthesis[47] for the production of a large variety of diverse patellamide-like cyclic peptides on a mg scale, the first two hypothetical roles for the DUFs, as described above, are the most important. As a consequence, the objective of this chapter is to characterise the C-terminal domains of unknown function, focusing on these first two hypothetical roles, to ultimately determinine whether the DUFs would constitute an essential part of a biotechnological 'toolkit' for the *in vitro* synthesis of cyanobactins. We turn to X-ray crystallography in an attempt to gain insight through structural characterisation, and isothermal titration calorimetry (ITC) and nuclear magnetic resonance spectroscopy (NMR) to investigate binding of the DUFs with plausible substrates: the precursor peptide and intermediates formed during biosynthesis. By expressing the DUFs alongside their associated domains, we can investigate what effect if any, the DUFs have on their activity.

## 2.2 Materials and Methods

N.B. References to PatG-DUF$_{sp.}$ and PatG-DUF$_{di.}$ described in this methods section refer to two different constructs of PatG-DUF with a slightly different amino acid sequence as discussed in section 2.3.1 and Fig. 2.7.

### 2.2.1 Expression and Purification of Native and SeMet PatG-DUF

The C-terminal domain of PatG from *Prochloron* sp. (PatG-DUF; amino acid residues 914-1186) was cloned from full-length *patG* into pEHISTEV plasmid with an N-terminal His$_6$ tag and a *Tobacco etch virus* (TEV) protease site[79]. A vector map can be seen in Fig. 2.4. The protein was expressed in *E. coli* BL21 (DE3) cells grown in Luria-Bertani (LB) medium supplemented with 50 μg mL$^{-1}$ kanamycin. Cultures were incubated at 37

°C, 200 rpm until $OD_{600}$ = 0.6. Cells were subsequently induced with 1 mM isopropyl-thiol-β-D-galactoside (IPTG) and further incubated at 18 °C, 200 rpm overnight. L-Selenomethionine-labelled (SeMet) PatG-$DUF_{sp.}$ was expressed from *E. coli* BL21 (DE3) cells grown in minimal medium supplemented with glucose-free nutrient mix (Molecular Dimensions), 50 μg $mL^{-1}$ kanamycin and 5 % glycerol. This medium was inoculated with overnight culture grown in LB, which had been washed three times in minimal medium. After 15 min growth at 37 °C, 60 mg $L^{-1}$ L-selenomethionine was added. The cultures were returned to 37 °C and incubated until $OD_{600}$ = 0.6, at which point 100 mg $L^{-1}$ lysine, phenylalanine, threonine and 50 mg $L^{-1}$ isoleucine and valine were added. After incubation for a further 20 min, expression was induced by the addition of 1 mM IPTG, and the cells were incubated at 18 °C for 24 h. For both native and SeMet variants, the cells were harvested by centrifugation (4000*g*, 4 °C, 15 min).

Cell pellets were resuspended in G-DUF lysis buffer (Appedix A) supplemented with 0.4 mg DNAse (Sigma-aldrich) per gram of wet cell pellet and cOmplete EDTA-free protease-inhibitor tablets (Roche; 1 per 50 mL resuspension). The cells were lysed via passage through a cell disruptor at 207 MPa (Constant Systems) and the cell debris was removed via centrifugation (40 000*g*, 4 °C, 20 min). The supernatant was loaded onto a pre-equilibrated Ni-Sepharose 6 Fast Flow column (GE Healthcare) at 4 °C, and the column was washed with G-DUF lysis buffer. The protein was eluted from the column in G-DUF elution buffer (Appendix A). The elution was passed over a desalt column (Desalt 16/10, GE Healthcare) in desalt buffer supplemented with 10 % glycerol (Appendix A). TEV-protease was added at a mass:mass ratio of 1 mg TEV to 10 mg of protein and the sample was incubated for 2 h at room temperature to remove the N-terminal $His_6$ tag. The digested sample was passed over a second Ni-Sepharose 6 Fast Flow column and the flow-through was collected and loaded onto a Highload 16/60 Superdex 200 gel-filtration column (GE Healthcare) equilibrated in gel filtration buffer supplemented in 10 % glycerol (Appendix A). The purity of the protein was analysed by SDS-PAGE, and its identity confirmed by LC-ESI-MS. Analysis of the mass spectrum of SeMet PatG-$DUF_{sp.}$ confirmed the successful incorporation of the five expected SeMet residues. Native and SeMet PatG-$DUF_{sp.}$ were concentrated to 4.75 mg $mL^{-1}$ for crystallography, using a 10 000 MWCO vivaspin centrifugal concentrator (Generon).

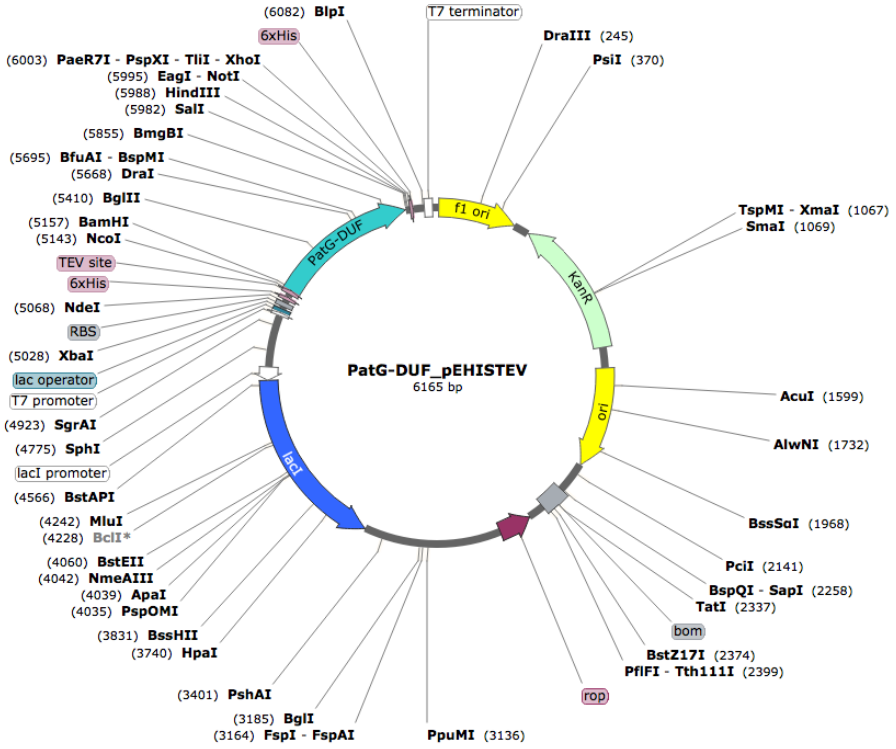



FIGURE 2.4: **Vector map of PatG-DUF**

### 2.2.2 Mutagenesis of PatG-DUF$_{sp.}$ to PatG-DUF$_{di.}$

Mutagenesis of PatG-DUF$_{sp.}$ was performed using established protocols[80].

### 2.2.3 CD Spectroscopy

Near-UV CD spectra were recorded on a Jasco J-18 spectro-polarimeter with PatG-DUF$_{sp.}$ and PatG-DUF$_{di.}$ samples in gel filtration buffer supplemented with 10 % glycerol.

### 2.2.4 Crystallography

Crystal screens were set up with the sitting-drop vapour diffusion method using a Gryphon robot (Art Robbins). The protein was screened against sparse-matrix screens composed of a range of known crystallisation conditions at 20 and 4 °C[81]. Diffraction-quality crystals were grown in a condition containing 0.04 M potassium phosphate, 16 % PEG 8000 and 20 % glycerol. Crystals of SeMet PatG-DUF$_{sp.}$ grew in the same condition

as native protein. A single SeMet PatG-DUF$_{sp.}$ crsystal was cryoprotected in mother liquor and flash-cooled in liquid nitrogen. A single wave-length anomalous dispersion (SAD) data set was collected at the Se *K* absorption edge at 100 K on beamline IO2 at Diamond Light Source. The structure was solved uising *AutoSol* and the chains were built into electron density using *AutoBuild* from the *PHENIX* crystallography suite[82]. The structure was manually rebuilt in *Coot*[83] and refined using *REFMAC5*[84]. TLS restraints were generated for refinement using the TLSMD server[85] and the structure was validated using *MolProbity*[86]. All sequence alignments were created using *Clustal Omega*[87] and presented using *ALINE*[88]. All structural images were generated using *PyMol*[89].

### 2.2.5 Expression and Purification of PatE′ and $^{15}$N-PatE′

For biochemical characterisation an engineered PatE variant was used, denoted PatE′. PatE′ has a single core peptide with the sequence ITACITFC (corresponding to the natural product patellamide D) and a C-terminal $His_6$ tag. The full PatE′ amino acid sequence is as follows: MDKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDA-GLEASKITACITFCAYDGELEHHHHHH. PatE′ was cloned into pBMS23CHIS by Dr Wael Houssen (University of Aberdeen). A vector map can be seen in Fig. 2.5.

PatE′ was expressed in *E. coli* BL21 (DE3) grown in LB medium supplemented with 100 μg mL$^{-1}$ ampicillin. Cultures were incubated at 37 °C, 200 rpm until $OD_{600}$ = 0.6. Cells were subsequently induced with 1 mM IPTG and further incubated at 37 °C, 200 rpm for 5 h. $^{15}$N-PatE′ was expressed in *E. coli* BL21 (DE3) cells grown in minimal medium prepared with $^{15}NH_4Cl$ (Sigma-aldrich) and supplemented with glucose-free nutrient mix (Molecular Dimensions), 100 μg mL$^{-1}$ ampicillin and 5 % glycerol. This medium was inoculated with overnight culture grown in LB and washed three times in minimal medium. The cultures were incubated at 37 °C until $OD_{600}$ = 0.6, induced with 0.5 mM IPTG and incubated at 30 °C overnight. Cells were harvested by centrifugation (4000*g*, 4 °C, 15 min).

Cell pellets were re-suspended in urea lysis buffer (Appendix A) and lysed by sonication at 15 microns (SoniPrep 150, MSE) with a pulse time of 30 s and a total on time of 4 min per 100 mL lysate solution. The lysate was cleared by centrifugation (40 000*g*, 20 °C, 20 min), and the supernatant loaded onto a Ni-Sepharose 6 Fast Flow column (GE

Healthcare). The column was washed with urea lysis buffer and peptide was eluted with urea elution buffer (Appendix A). The peptide elution was incubated in the presence of 10 mM dithiothreitol (DTT) for 2 h prior to passage down a Highload 16/60 Superdex 75 gel filtration column (Ge Healthcare) pre-equilibrated in gel filtration buffer. The purity of the peptide was analysed by SDS-PAGE, and its identity confirmed by MALDI-TOF-MS.

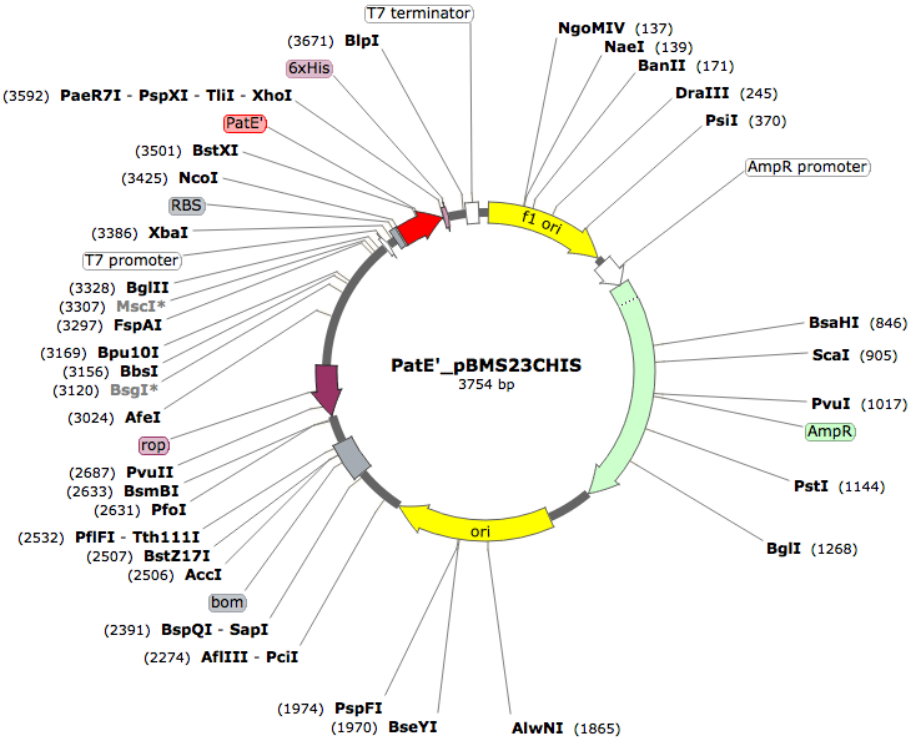



FIGURE 2.5: **Vector map of PatE′**

### 2.2.6 Expression and Purification of PatD

Codon optimised full-length PatD with an N-terminal thrombin-cleavable $His_6$ tag was purchased from DNA2.0 in the pJexpress 411 vector. It was expressed from *E. coli* BL21 (DE3) cells, grown in autoinduction medium at 20 °C for 48 h[90]. Cells were harvested by centrifugation (4000*g*, 4 °C, 15 min). Cell pellets were resuspended in general lysis buffer (Appedix A) supplemented with 0.4 mg DNAse (Sigma-aldrich) per gram of wet cell pellets and cOmplete EDTA-free protease-inhibitor tablets (Roche; 1 per 50 mL resuspension). The cells were lysed via passage through a cell disruptor at 207 MPa (Constant Systems) and clarified by centrifugation (40 000*g*, 4 °C, 20 min). Cleared

lysate was applied to Ni-Sepharose 6 Fast Flow column (GE Healthcare) equilibrated in general lysis buffer. The protein was washed with general lysis buffer, then eluted using general elution buffer (Appendix A). The elution was then passed over a Highload 16/60 Superdex 200 gel filtration column (GE Healthcare) equilibrated in gel filtration buffer.

### 2.2.7 Heterocyclisation of $^{15}$N-PatE′ using PatD

Pure $^{15}$N-PatE′ (100 μM) was incubated in the presence of PatD (5 μM), ATP (5 mM), and $MgCl_2$ (5 mM) in gel filtration buffer at 27 °C overnight to ensure complete heterocyclisation. Following incubation, heterocyclised $^{15}$N-PatE′ was isolated via size exclusion chromatography by passage down a Highload 16/60 Superdex 75 gel filtration column (GE Healthcare) equilibrated in gel filtration buffer. The reaction was monitored by MALDI-TOF-MS; a loss of 18 Da corresponding to the loss of water, was observed for each heterocycle formed.

### 2.2.8 ITC Data Collection and Analysis

ITC experiments were performed using a VP-ITC instrument (MicroCal) in gel filtration buffer at 20 °C. Both PatE′ and PatG-DUF$_{di.}$ were dialysed into freshly prepared gel filtration buffer for 2 days (changing to fresh buffer after the first day) at 4 °C to ensure that the buffers for the experiment were matched. A cell solution of 40 μM PatG-DUF$_{di.}$ and a syringe solution of 600 μM PatE′ were prepared (using the dialysis buffer to dilute the samples) and degassed at 18 °C for 15 min. The PatE′ solution was titrated into the PatG-DUF$_{di.}$ solution as follows: an initial injection of 2 μL followed by injections of 5 μL at a speed of 0.5 μL min$^{-1}$ with a delay of 4 min between injections. The syringe was stirred at 307 rpm for the duration of the experiment. The raw data were processed using the MicroCal *Origin* software. The baseline was adjusted and the integrations were performed manually.

### 2.2.9 NMR Binding Experiments

NMR experiments were performed at 10 °C in a Bruker AVANCE III 500 MHz spectrometer equipped with a 5 mm TXIz probe. The instrument was run using the *TopSpin* software (Bruker). The sample for binding experiments consisted of 100

μM $^{15}$N-PatE′ in gel filtration buffer supplemented with 5 % $D_2O$, and a $^{1}$H-$^{15}$N HSQC spectrum was recorded. Aliquots of 1.6 mM PatG-$DUF_{di.}$ were added to final concentrations of 50, 100 and 200 μM and HSQCs were recorded after each addition. The spectra were overlaid using *TopSpin* to identify changes (if any) in the spectra subsequent to addition of PatG-$DUF_{di.}$. The $^{1}$H-$^{15}$N HSQC sepctra were acquired with Water-gate supression[91] at 1024 x 128 points and a digital resolution of 9.8 and 19.0 Hz for the $^{1}$H and $^{15}$N dimensions respectively. The experiments were repeated using a modified PatE′ containing four heterocycles under the same conditions. This resulted in a poorer signal-to-noise ratio and therefore the experiment was repeated using 250 μM modified $^{15}$N-PatE′ and aliquots of 2 mM PatG-$DUF_{di.}$ were added to final concentrations of 125, 250 and 500 μM.

### 2.2.10 Expression and Purification of PatGmacDUF

PatGmacDUF (amino acid residues 515 - 1191) was cloned into pEHISTEV plasmid by Dr Jesko Koehnke. A vector map can be seen in Fig. 2.6. The protein was expressed in *E. coli* BL21 (DE3) cells grown in LB medium supplemented with 50 μg $mL^{-1}$ kanamycin. Cultures were incubated at 37 °C, 200 rpm until $OD_{600}$ = 0.6. Cells were subsequently induced with 1 mM IPTG and further incubated at 18 °C, 200 rpm overnight. Cells were harvested by centrifugation (4000*g*, 4 °C, 15 min).

Cell pellets were resuspended in general lysis buffer (Appedix A), supplemented with 0.4 mg DNAse (Sigma-aldrich) per gram of wet cell pellet and cOmplete EDTA-free protease-inhibitor tablets (Roche; 1 per 50 mL resuspension). The cells were lysed via passage through a cell disruptor at 207 MPa (Constant Systems) and the cell debris was removed via centrifugation (40 000*g*, 4 °C, 20 min). The supernatant was loaded onto a pre-equilibrated Ni-Sepharose 6 Fast Flow column (GE Healthcare) at 4 °C, and the column was washed with general lysis buffer. The protein was eluted from the column in general elution buffer (Appendix A). The elution was passed over a desalt column (Desalt 16/10, GE Healthcare) in desalt buffer 2 (Appendix A). TEV-protease was added at a mass:mass ratio of 1 mg TEV to 10 mg of protein and the sample was incubated for 2 h at room temperature to remove the N-terminal $His_6$ tag. The digested sample was passed over a second Ni-Sepharose 6 Fast Flow column and the flow-through was loaded onto a MonoQ column (GE Healthcare) equilibrated in desalt buffer 2. Protein

was eluted from the MonoQ column through a linear NaCl gradient. Finally the protein was passed down a Highload 16/60 Superdex 200 gel-filtration column (GE Healthcare) equilibrated in gel filtration buffer supplemented in 10 % glycerol (Appendix A).

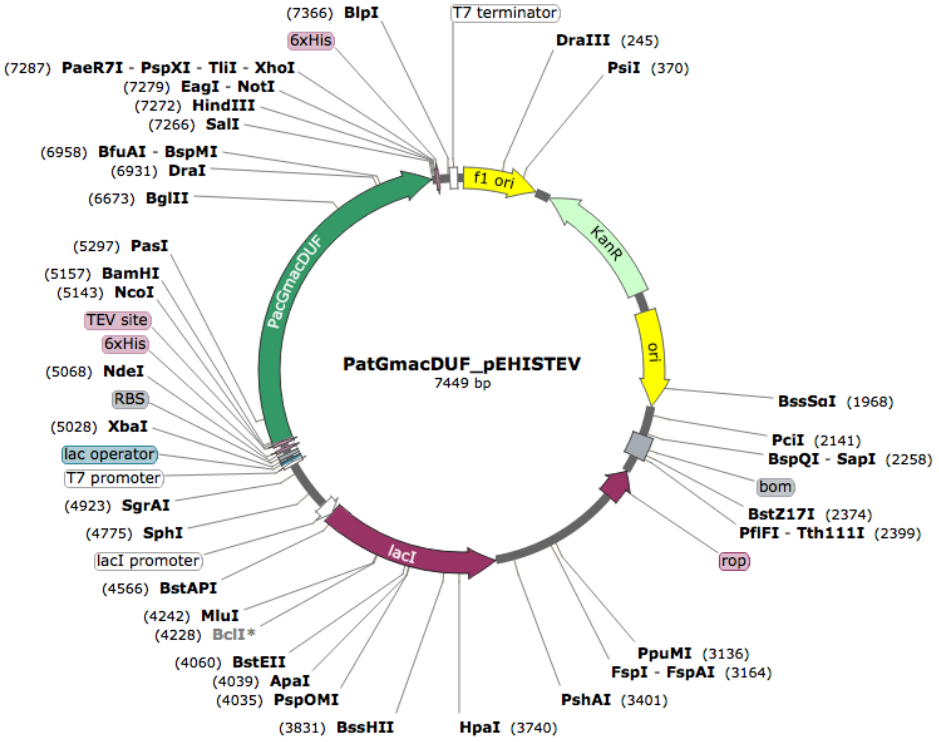



FIGURE 2.6: **Vector map of PatGmacDUF**

### 2.2.11 Expression and Purification of PatGmac

PatGmac was expressed from pEHISTEV vector and purified as described (section 5.5.2).

### 2.2.12 Measuring Macrocyclisation Rate by HPLC

$IT^{MeOx}AC^{ThH}IT^{MeOx}AC^{ThH}$AYDGE substrate peptide was generated as described (section 5.2.4) and purified via passage down a Highload 16/60 Superdex 30 gel filtration column (GE Healthcare). Macrocyclisation of pure $IT^{MeOx}AC^{ThH}IT^{MeOx}AC^{ThH}$AYDGE was carried out as described (section 5.2.4). Reaction progress was monitored using an Agilent infinity 1260 series RP-HPLC equipped with a MWD detector and a single quadropole MS using a Nucleodur $C_4$ (Macherey-Nagel) column (10 µm x 4.6 x 250 mm). 50 µL samples were taken at 0, 2, 4, 6, 8, 10, 12, 24, 48 and 72 h. Samples were

eluted from the column using an aqueous 5 mM ammonium carbonate - MeCN gradient with a flow rate of 1 mL $min^{-1}$ as follows: 0 - 3 min 5 % MeCN, 3 - 20 min 5 - 95 % MeCN, 20 - 22 min 95 % MeCN, 22 - 23 min 95 - 5 % MeCN.

## 2.3 Results and Discussion

### 2.3.1 Cloning, Expression and Purification of PatG-DUF

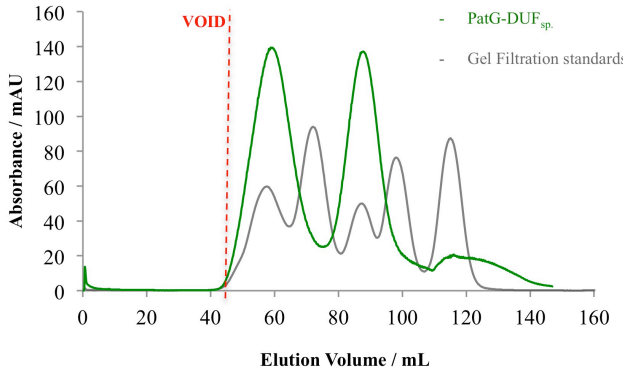
Due to the high sequence homology between the C-terminal domains of unknown function of PatA and PatG (Fig. 2.1), the decision to initially focus on PatG-DUF was arbitrary. In an attempt to isolate and subsequently characterise the domain, a truncated construct (amino acid residues 914-1186) of the C-terminus was designed based on sequence homology (figure 2.3) and cloned into pEHISTEV vector. This construct (PatG-DUF) was cloned from gDNA isolated from *Prochloron sp.* that were previously isolated from Pacific reef samples. The exact strain could not be determined and the amino acid sequence of this PatG-DUF construct (hereafter referred to as PatG-$DUF_{sp.}$) differs at three points (E958D, R959Y and V1037M) from the deposited amino acid sequence of PatG isolated from *Prochloron didemni* (accession No. AAY21556.1; Fig. 2.7). It is unclear as to whether these differences arose as a result of natural variance amongst species or a cloning artefact.

| | | |
|---|---|---|
| PatG-$DUF_{sp.}$ | V E A S T A F S G N V Y A L G T I G Y D F G D E A R R D T F K E R M A D P Y D | 952 |
| PatG-$DUF_{di.}$ | V E A S T A F S G N V Y A L G T I G Y D F G D E A R R D T F K E R M A D P Y D | 952 |
| PatG-$DUF_{sp.}$ | A R Q M V E R L D R N P D E A R S L I W T L N L E G D V I Y A L D P K G P F A | 991 |
| PatG-$DUF_{di.}$ | A R Q M V D Y L D R N P D E A R S L I W T L N L E G D V I Y A L D P K G P F A | 991 |
| PatG-$DUF_{sp.}$ | T N V Y E I F L Q M L A G Q L E P E T S A D F I E R L S V P A R R T T R T V E | 1030 |
| PatG-$DUF_{di.}$ | T N V Y E I F L Q M L A G Q L E P E T S A D F I E R L S V P A R R T T R T V E | 1030 |
| PatG-$DUF_{sp.}$ | L F S G E V V P V V N V R D P R G M Y G W N V N A L V D A A L A T V E Y E E A | 1069 |
| PatG-$DUF_{di.}$ | L F S G E V M P V V N V R D P R G M Y G W N V N A L V D A A L A T V E Y E E A | 1069 |
| PatG-$DUF_{sp.}$ | D E D S L R Q G L T A F L N R V Y H D L H N L G Q T S R D R A L N F T V T N T | 1108 |
| PatG-$DUF_{di.}$ | D E D S L R Q G L T A F L N R V Y H D L H N L G Q T S R D R A L N F T V T N T | 1108 |
| PatG-$DUF_{sp.}$ | F Q A A S T F A Q A I A S G R Q L D T I E V N K S P Y C R L N S D C W D V L L | 1147 |
| PatG-$DUF_{di.}$ | F Q A A S T F A Q A I A S G R Q L D T I E V N K S P Y C R L N S D C W D V L L | 1147 |
| PatG-$DUF_{sp.}$ | T F Y D P E H G R R S R R V F R F T L D V V Y V L P V T V G S I K S W S L P G | 1186 |
| PatG-$DUF_{di.}$ | T F Y D P E H G R R S R R V F R F T L D V V Y V L P V T V G S I K S W S L P G | 1186 |

FIGURE 2.7: **Sequence alignment of PatG-DUF cloned from DNA isolated from *Procholoron sp.* (PatG-$DUF_{sp.}$) and PatG-DUF isolated from *P. didemni* (PatG-$DUF_{di.}$).** Sequence differences are highlighted in green.

Consequently PatG-DUF$_{sp.}$ was modified by site-directed mutagenesis to generate PatG-DUF with the sequence from *P. didemni* (hereafter referred to as PatG-DUF$_{di.}$). Any protocols, results and discussion that are applicable to both PatG-DUF$_{sp.}$ (E958, R959 and V1037) and PatG-DUF$_{di.}$ (D958, Y959 and M1037) constructs will still be referred to more generally, as PatG-DUF.

PatG-DUF, subsequent to iterative rounds of optimisation of the expression and purification conditions (data not shown) was expressed and purified as described (section 2.2.1 and 2.2.2). Both PatG-DUF$_{sp.}$ and PatG-DUF$_{di.}$ behaved in an identical fashion, eluting from the gel filtration column as two separate peaks (Fig. 2.8). The first peak, eluting at approximately 55 mL corresponds with a high order multimer (>600 kDa) and was not expected to be biologically relevant. The second peak, eluting at approximately 90 mL contains pure PatG-DUF, as determined by SDS-PAGE and LC-ESI-MS, yielding 5 mg $L^{-1}$ culture. Comparison with standards (Bio-Rad) confirm that PatG-DUF elutes from the gel filtration column as a monomer (Fig. 2.8).

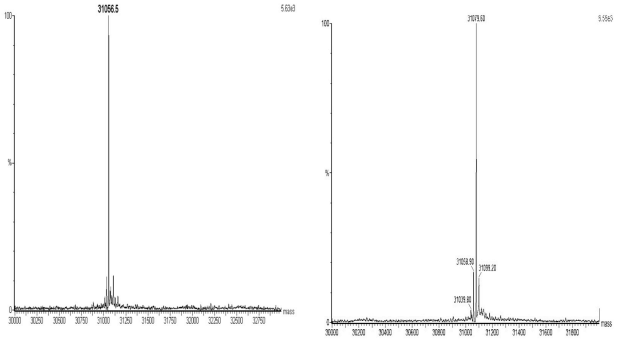



(a)

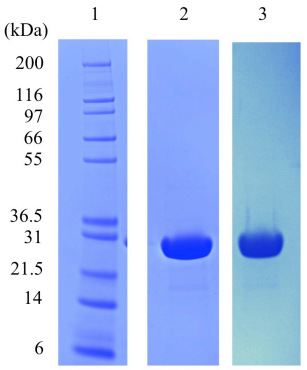
**Gel Filtration Standards**

| Elution Volume \ mL | Protein | Mass / Da |
|---|---|---|
| 57 | Thyroglobulin | 670 000 |
| 62 | γ-Globulin | 158 000 |
| 87 | Ovalbumin | 44 000 |
| 98 | Myoglobin | 17 000 |
| 115 | Vitamin $B_{12}$ | 1 350 |

(b)



(c) (d)




(e)

FIGURE 2.8: **PatG-DUF purification** (a) Highload 16/60 Superdex 200 gel filtration chromatogram of PatG-DUF$_{sp.}$ (green) and Bio-Rad gel filtration standards (grey); (b) elution volume and molecular weight of gel filtration standards (Bio-Rad); (c) LC-ESI-MS of PatG-DUF$_{sp.}$ processed to 0.1 Da; (d) LC-ESI-MS of PatG-DUF$_{di.}$ processed to 0.1 Da; (e) SDS-PAGE of PatG-DUF$_{sp.}$ (lane 2) and PatG-DUF$_{di.}$ (lane 3) compared with Mark12$^{TM}$ molecular weight marker (lane 1).

Both PatG-DUF$_{sp.}$ and PatG-DUF$_{di.}$ were dialysed into gel filtration buffer supplemented with 10 % glycerol and a Near-UV CD spectra were recorded (Fig. 2.9). These spectra revealed no significant difference in the overall tertiary structure between the two variants and so we concluded that either sequence would be suitable for structural characterisation.

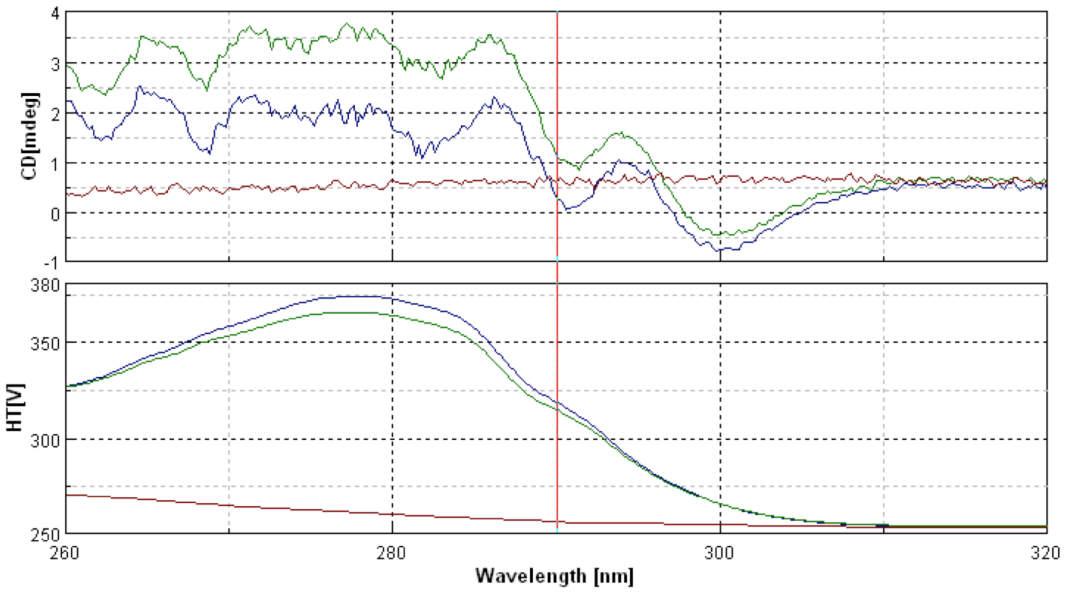


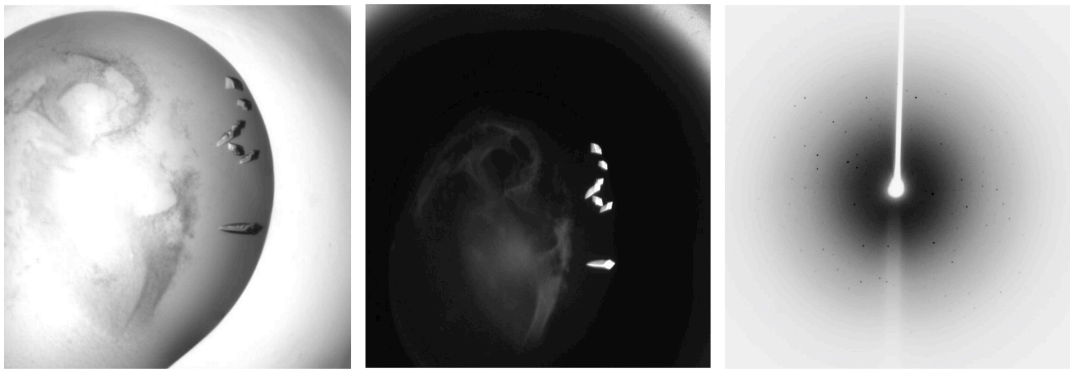

FIGURE 2.9: **Near UV CD-spectra of PatG-DUF$_{sp.}$ and PatG-DUF$_{di.}$** PatG-DUF$_{sp.}$ (green) and PatG-DUF$_{di.}$ (blue) are overlaid. A blank containing only buffer is shown in red. The CD-spectra overlay well, suggesting the sequence variations (three-point mutations) do not alter the tertiary structure of the domain.

Crystallography was carried out on PatG-DUF$_{sp.}$ as it had been purified first, and gave crystals. We carried out biophysical and biochemical studies with PatG-DUF$_{di.}$ as we felt these data would be more informative, with a known database match.

### 2.3.2 Crystallography of PatG-DUF$_{sp.}$

Pure PatG-DUF$_{sp.}$ at 4.75 mg mL$^{-1}$ was subjected to crystallisation trials using a range of sparse-matrix screens[81]. After three days crystals were formed in a precipitant composition of 0.04 M potassium phosphate, 6 % PEG 8000 and 20 % glycerol (Fig. 2.10). A single crystal was picked and a native data set was collected in-house to 2.1 Å (Fig. 2.10, Table 2.1).

Determination of electron density, and thus the crystallographic structure of the protein from X-ray diffraction data requires knowledge of the amplitudes and the phases of the scattered waves[92]. While the amplitudes can be calculated from the intensity of the reflections in the diffraction pattern, the phase information is lost in the experiment[93]. If a structurally similar model exists, initial phases can be estimated by placing the



(a) (b) (c)

FIGURE 2.10: **PatG-DUF$_{sp.}$ crystallography** (a) Image of PatG-DUF$_{sp.}$ crystals taken from Rigaku Minstrel imager. (b) UV image of crystals taken from Rigaku Minstrel imager. (c) Single crystal diffraction to 2.1 Å recorded in house.

TABLE 2.1: **Data collection statistics for PatG-DUF$_{sp.}$ native data set**. Statistics are average values; values for the highest resolution shell are included in parentheses.

| **PatG-DUF$_{sp.}$ native data set** | |
|---|---|
| Wavelength (Å) | 1.54 |
| Space group | $P2_1\ 2_1\ 2$ |
| Unit-cell parameters (Å, °) | a = 64.2, b = 95.5, c = 40.3 |
| | $\alpha = \beta = \gamma = 90.0$ |
| Resolution (Å) | 38.3-2.10 (2.16-2.1) |
| I/σI | 12.9 (4.3) |
| $R_{merge}$ (%) | 8.8 (36.8) |
| Completeness (%) | 99.3 (95.7) |
| Multiplicity | 5.0 (4.2) |

model of the known protein into the unit cell of the new protein crystal, in a process called molecular replacement[92]. However in the case of PatG-DUF$_{sp.}$, the structure has not been determined for any of its sequence homologues, so no such model exists. Therefore the structure cannot be determined using the native dataset already collected. Phase information had to be obtained experimentally, in this case by the use of single wavelength anomalous dispersion (SAD).

Different elements absorb X-rays at specific wavelengths. This absorption drops

sharply at wavelengths just bellow their emission wavelengths, a feature known as the element's absorption edge. If the wavelength of the incident X-ray is close to an element's absorption edge, the element will exhibit anomalous dispersion[92]. Anomalous dispersion (sometimes called anomalous scatter) results in a break down in Friedel's law, and so reflections *hkl* are no longer equal in intensity to reflections *-h-k-l*[92]. This difference between Friedel pairs can be used to locate the atom responsible for the anomalous scattering within the unit cell, providing phase information for this atom, which can be used to calculate the phases of the reflections of the protein[92,93].

The absorption edge of light atoms, such as carbon, nitrogen and oxygen is away from the wavelength of X-rays used in crystallography, and so they do not contribute to anomalous dispersion within a protein unit cell[85]. However the absorption edge of heavier atoms is within this range, and their incorporation into a unit cell would provide the anomalous scattering necessary to solve the crystal structure[92,93].

The most common and efficient method to introduce heavy atoms into the protein crystal is to replace all the methionine residues with selenomethionine (SeMet), an analogue containing a selenium atom in place of the sulfur. This provides an isomorphous crystal with exploitable anomalous dispersion characteristics. This can be achieved by the expression of protein in *E. coli* in a minimal medium where the L-methionine is entirely replaced with L-selenomethionine. Inhibition of methionine biosynthesis by adding large concentrations of leucine, isoleucine, phenylalanine, lysine and threonine shortly before induction enhances the chance of complete selenomethionine incorporation during protein expression[94].

Purification of SeMet PatG-DUF$_{sp.}$ was identical to that of native protein. However as is anecdotally typical with SeMet incorporation, a lower yield of 0.46 mg L$^{-1}$ of culture was observed. LC-ESI-MS analysis confirmed all five selenium residues in PatG-DUF$_{sp.}$ had been successfully replaced with SeMet (Fig. 2.11).

The lower peak at 30 758.7 Da represents an impurity, which has not yet been identified. However no significant impurity was observed by SDS-PAGE analysis (Fig. 2.11) and the impurity did not affect crystalisation. SeMet PatG-DUF$_{sp.}$ was concentrated to 4.75 mg L$^{-1}$ and screened for crystallisation. After three days crystals were observed in the same precipitant conditions as the native protein. The largest crystal was picked,

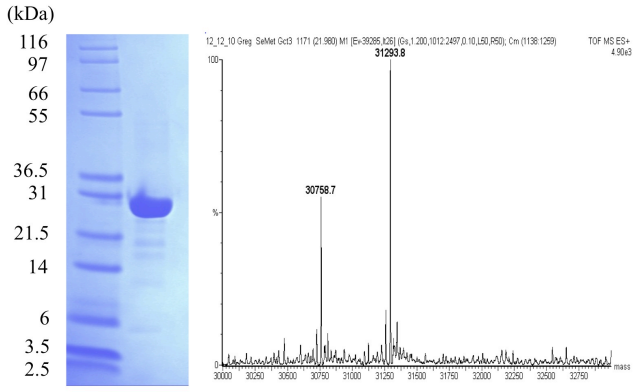



FIGURE 2.11: **SeMet PatG-DUF$_{sp.}$ production** (a) SDS PAGE analysis showing purified SeMet PatG-DUF$_{sp.}$. (b) LC-ESI-MS of SeMet PatG-DUF$_{sp.}$ (Mass = 31 293.8 Da). A mass increase compared to the mass of the native protein (31 056.5 Da) corresponds with incorporation of all five possible selenomethionine residues. The smaller peak (30758.7 Da) corresponds to an impurity that hasn't presently been identified.

which diffracted to 1.72 Å under synchrotron radiation. Data collection and refinement statistics can be seen in Table 2.2.

The SeMet PatG-DUF$_{sp.}$ crystal belonged to the space group $P2_12_12_1$ and contained one monomer in the asymmetric unit. The structure has a fold comprising of seven α-helices and thirteen β-strands. Analysis with secondary structure matching at the European Bioinformatics Institute and *TopSearch* both suggest the fold is novel, with the closest matches being a series of unrelated enterotoxins, giving r.m.s.d. values of between 3 and 4 Å over ∼80 residues [95,96].

The refined model contains residues 920-1064 and 1077-1184 (residue numbering corresponds to full-length PatG). The missing residues are at the N- and C-termini of the model and in a connecting loop between α-helices 5 and 6, and are presumed to be disordered. The model includes a single disulfide bond between C1136 and C1142 (Fig. 2.12). Regarding the sequence differences between PatG-DUF$_{sp.}$ and PatG-DUF$_{di.}$ (Fig. 2.7): E958 and R959 are located on the surface, whilst V1037 is located on the edge of a hydrophobic pocket near the surface (Fig. 2.12). None of these sequence alterations would seem likely to alter the structure, which is consistent with the near-UV CD spectra (Fig. 2.9).

Examination of the identified anomalous scattering atoms revealed the Se atoms of the four ordered methionine residues (the fifth selenomethionine at the start of protein sequence is not visible in the model and presumed disordered). However, an additional anomalous signal was observed at the interface of one monomer and its symmetry mate. This signal corresponded with a significant sphere of unaccounted-for electron density, which was coordinated in a tetrahedral arrangement by residues E977, H1087, H1090 and D1088* from the symmetry mate (PatG-DUF*$_{sp.}$) creating a dimer. Based on the anomalous signal, coupled with this coordination, the electron density was attributed to a $Zn^{2+}$ ion. Since zinc was not deliberately added to any of the purification buffers, or crystallisation conditions, it is presumed $Zn^{2+}$ is bound in the cell. An equivalent $Zn^{2+}$ binding site in the symmetry mate was also observed (Fig. 2.12).

TABLE 2.2: **Data collection and refinement statistics for SeMet PatG-DUF$_{sp}$.** An anomalous data set was collected using a single crystal on a beamline I02 at the Diamond Light Source. Statistics are average values; values for the highest resolution shell are included in parentheses.

| **SAD data set** | |
|---|---|
| Wavelength (Å) | 0.9797 |
| Space group | $P2_1\ 2_1\ 2$ |
| Unit-cell parameters (Å, °) | a = 64.1, b = 95.2, c = 40.2 |
| | $\alpha = \beta = \gamma = 90.0$ |
| Resolution (Å) | 53.2-1.7 (1.8-1.7) |
| I/σI | 12.1 (3.4) |
| $R_{merge}$ (%) | 14.6 (89.6) |
| Completeness (%) | 99.9 (100) |
| Multiplicity | 14.0 (11.3) |
| Anomalous completeness | 99.9 (99.9) |
| Anomalous multiplicity | 7.3 (5.7) |
| **Refinement** | |
| R-factor (%) | 17.5 (23.9) |
| $R_{free}$ (%) | 20.0 (29.5) |
| R.m.s.d., bond lengths (Å) | 0.005 |
| R.m.s.d., bond angles (°) | 0.932 |
| **No. of non-H atoms** | |
| Protein atoms | 2109 |
| Solvent atoms | 269 |
| Heterogen atoms | 1 |
| ***B* factors Å$^2$** | |
| All | 26.0 |
| Protein | 25.1 |
| Ligand ($Zn^{2+}$) | 15.9 |
| Water | 32.8 |

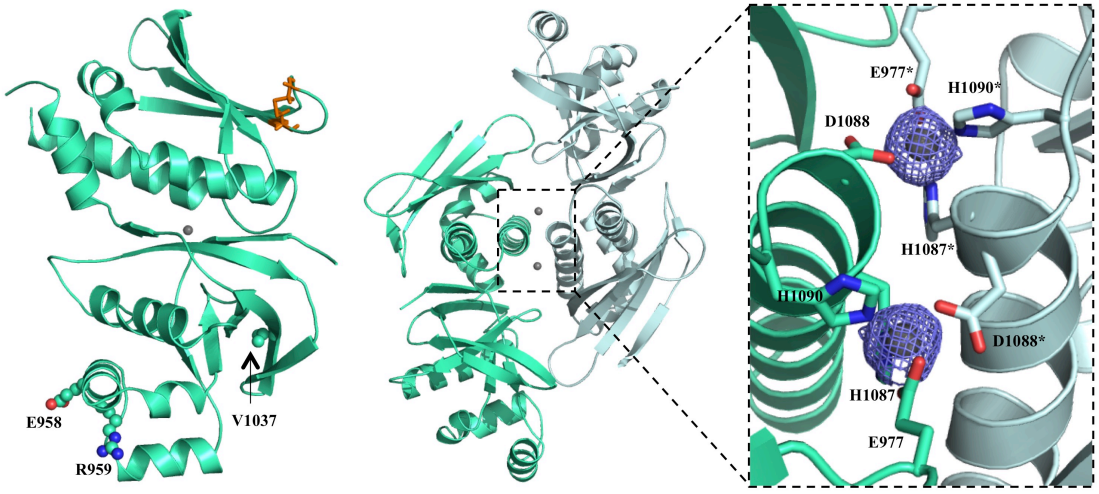



FIGURE 2.12: **X-ray crystal structure of PatG-DUF$_{sp.}$.** (a) PatG-DUF$_{sp.}$ represented as a cartoon. The disulfide bond between C1136 and C1142 is shown as orange sticks and the $Zn^{2+}$ is shown as a grey sphere. PatG-DUF$_{sp.}$ amino acids which differ from those in PatG-DUF$_{di.}$ are highlighted as spheres for clarity. (b) X-ray crystal structure of the PatG-DUF$_{sp.}$ dimer represented as a cartoon. (c) Enlargement of the $Zn^{2+}$ coordination site. E977, H1087, D1088 and H1090 are shown as green sticks, and E977*, H1087*, D1088* and H1090* are shown as cyan sticks. Difference electron density ($F_o$-$F_c$) contoured at 3σ with phases calculated from a model that was refined without $Zn^{2+}$ present is shown as a blue isomesh.

Analysis with *PISA* - software that enables the exploration of macromolecular interfaces[97] reveals that there are almost no contacts between the amino acid residues at the interface, and that removing the $Zn^{2+}$ ions *in silico* reduces the complex formation significance score to 0.056, indicating $Zn^{2+}$ is crucial in mediating dimer interactions. Since $Zn^{2+}$ was not added externally, its presence in the crystal could suggest that the dimer is present in solution. However PatG-DUF eluted from the gel filtration column as a monomer (Fig. 2.8), preventing a firm assignment of dimer *versus* monomer. Interestingly the residues involved in $Zn^{2+}$ binding are not all conserved in PatA-DUF (Fig. 2.13). While the H1087Y substitution might still coordinate $Zn^{2+}$, the H1090R substitution would not. Additionally these residues are not well conserved in the majority of PatG-DUF homologues (with notable exceptions being TruG-DUF and ArtG-DUF) or in any of the PatA-DUF homologues (Fig. 2.13). Consequently we cannot make any claims regarding the biological relevance of a dimeric assembly for the DUF domain.

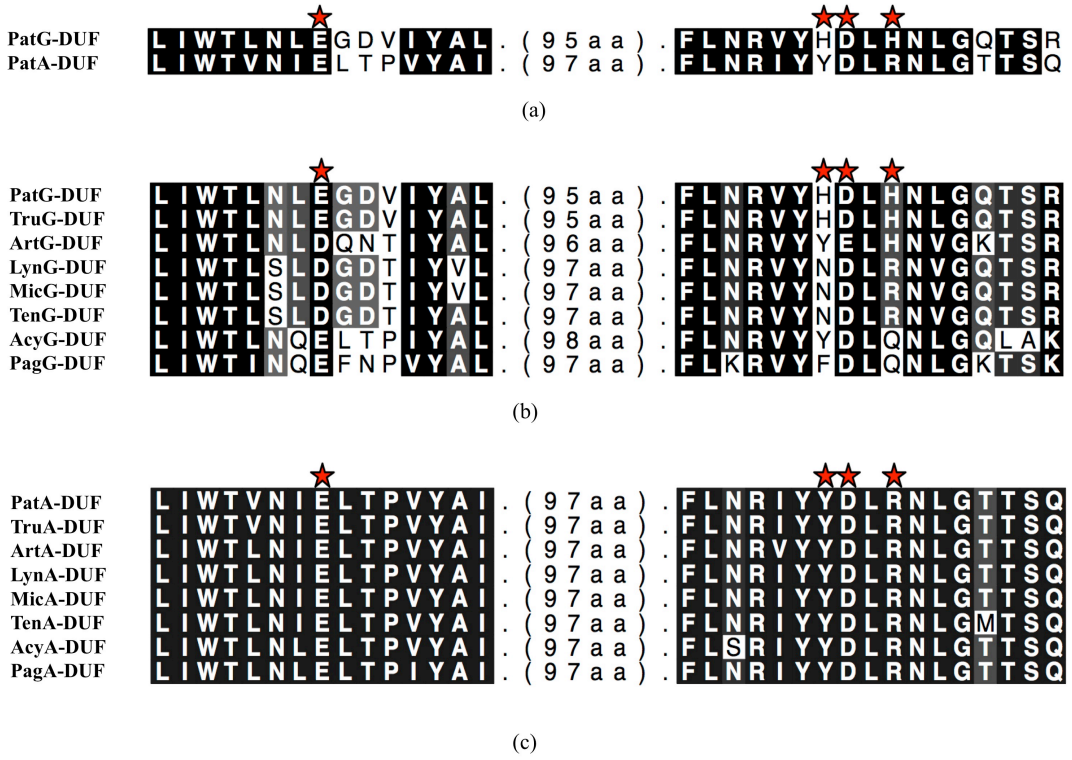



FIGURE 2.13: **Truncated DUF sequence alignments** (a) PatG-DUF and PatA-DUF, (b) PatG-DUF and its homologues, (c) PatA-DUF and its homologues. The residues involved in $Zn^{2+}$ coordination in PatG-DUF are highlighted with red stars.

### 2.3.3 Cloning, Expression and Purification of PatA-DUF

In an attempt to isolate stable, homogenous PatA-DUF for structural characterisation, a construct of the C-terminal domain based on the PatG-DUF construct that had already been used in structural studies was designed (figure 2.1), and cloned into pEHISTEV plasmid. As a set of preliminary expression experiments, PatA-DUF was expressed from a set of 10 mL *E. coli* BL21 (DE3) cultures in LB, incubated until $OD_{600}$ = 0.6 and induced with 0.1, 0.5 or 1 mM IPTG and expressed at 15, 25 or 37 °C overnight. This small-scale expression test confirmed induction with 0.1 mM IPTG at 15 °C yields the greatest amount of protein (Fig. 2.14).

However in these experiments the majority of the protein did not bind the Ni beads, with a large amount of the sample present in the flow-through (Fig. 2.14). This might be because the protein is aggregated in solution and so the $His_6$ tag is unaccessible to the Ni beads. A variety of buffer conditions were tested, but none resulted in a noticeable improvement (data not shown). Therefore PatA-DUF was cloned into

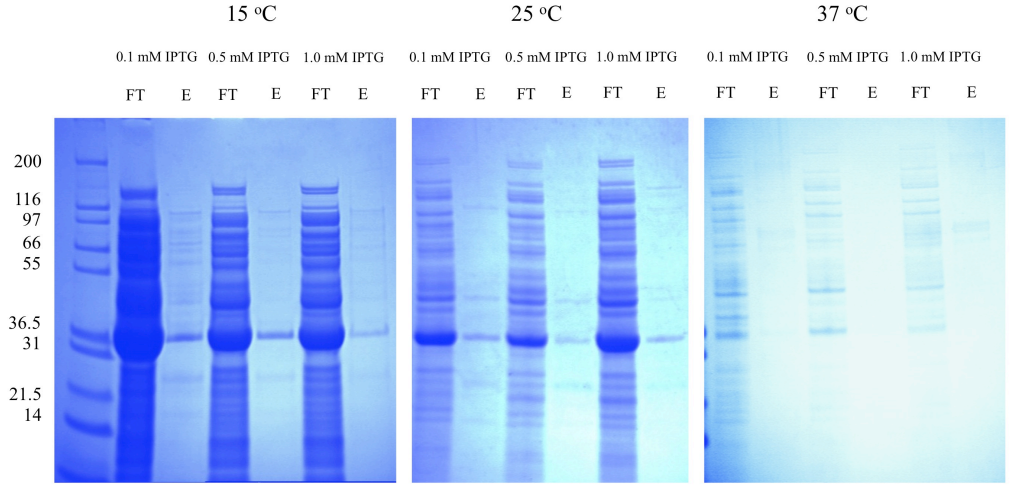



FIGURE 2.14: **PatA-DUF expression** SDS-PAGE of PatA-DUF expression test: Flowthrough and Elution from Ni column loaded with lysate from *E. coli* BL21 cells induced with 0.1, 0.5 and 1.0 mM IPTG and incubated at 15, 25 and 37 °C overnight.

pEHISSUMOTEV, a vector containing a TEV-cleavable Small Ubiquitin-like Modifier (SUMO) tag preceded directly at its N-terminus by a $His_6$ tag. SUMO tags have been shown to enhance protein expression, and promote correct protein folding[98]. Additionally, the covalent attachment of the ∼12 kDa tag should sufficiently extend the $His_6$ tag from the core of the protein to promote binding to the Ni column. Initial signs were promising with yields >100 mg $L^{-1}$ of culture eluting from the Ni column. However, once the tag was removed the protein eluted from the gel filtration column in the void volume, implying the sample was entirely aggregated (Fig. 2.15).

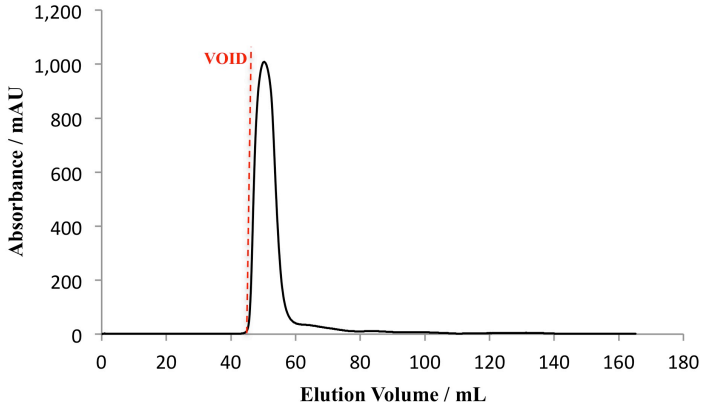



FIGURE 2.15: **PatA-DUF purification** Gel Filtration chromatograph of PatA-DUF after removal of the SUMO tag. PatA-DUF elutes as a single peak in the void volume inplying the entire protein sample is aggregated.

At present, all attempts to purify PatA-DUF have been unsuccessful. Consequently, given the high homology between the C-terminal domains of PatA and PatG, all subsequent attempts to identify a role for the domains have focused on the biochemical and biophysical characterisation of PatG-DUF.

### 2.3.4 Investigating Potential Binding Partners for PatG-DUF$_{di.}$

The X-ray crystal structure for PatG-DUF$_{sp.}$ is novel, and so provides no immediate insight as to why the DUFs are conserved throughout cyanobactin biosynthetic pathways. In an attempt to elucidate the function of this domain, and importantly, determine whether it has a direct role in patellamide biosynthesis, a number of potential binding partners for PatG-DUF were explored. Obvious targets to test include the patellamide precursor peptide PatE, the modified precursor biosynthetic intermediates, and the final macrocyclic product. If binding is observed, subsequent experiments can be designed to test hypotheses regarding the function of the domain. A modified precursor peptide PatE′ was used in place of the natural sequence.

Initial binding experiments were performed using ITC. Titrating PatE′ into a solution of PatG-DUF$_{di.}$ reproducibly results in a weak, but detectible, titration curve, indicative of a binding interaction (Fig. 2.16). However analysis of the data yields unlikely stoichiometries of around ten molecules of PatG-DUF$_{di.}$ binding to one molecule of PatE′. Furthermore, close inspection of the cell solution revealed heavy precipitation at the end of the experiment. Therefore it seems likely that the exothermic signal observed arises from protein aggregation rather than a genuine binding interaction between PatE′ and PatG-DUF$_{di.}$.

To resolve the ambiguity of the ITC data, binding experiments between PatG-DUF$_{di.}$ and PatE′ were performed using NMR spectroscopy. Using $^{15}$N labelled PatE′ a $^{1}$H-$^{15}$N-HSQC spectrum was recorded in the absence, and presence of 0.5, 1.0, and 2.0 molar equivalents of PatG-DUF$_{di.}$ (Fig. 2.17), where a binding event would give rise to significant changes in the chemical shift of the N-H cross peaks.

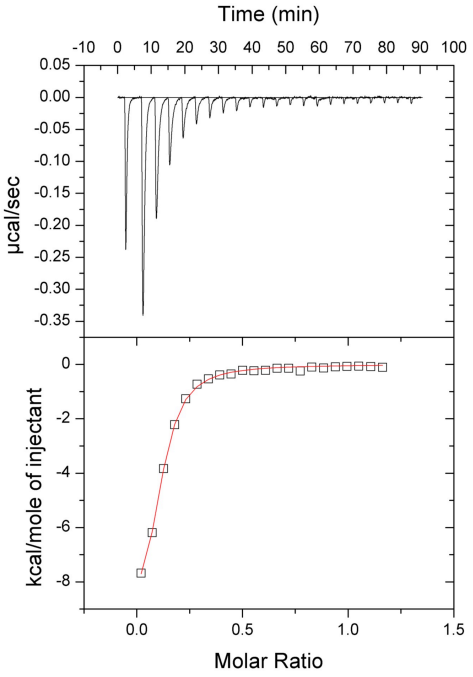



FIGURE 2.16: **ITC of PatG-DUF$_{di.}$ and PatE′** ITC data of PatE′ titrated into PatG-DUF$_{di.}$ solution. The top panel shows raw data representing heat evolved in response to injections, the bottom panel shows the integrated heats of injections (□) and the best fit (-) to the one-site model (*Origin*).

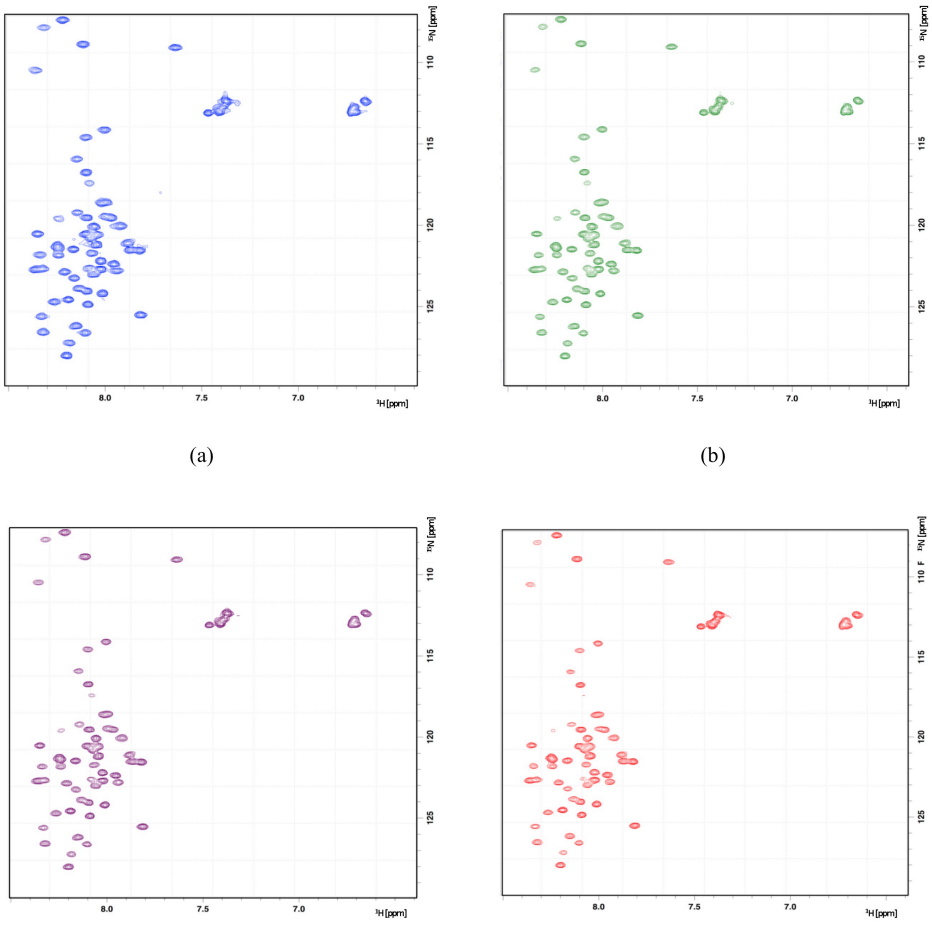



FIGURE 2.17: **NMR of $^{15}$N-PatE′ + PatG-DUF$_{di}$** $^{1}$H-$^{15}$N-HSQC of $^{15}$N-PatE′ (blue), with 0.5 (green), 1.0 (purple) and 2.0 (red) equivalents of PatG-DUF$_{di.}$.

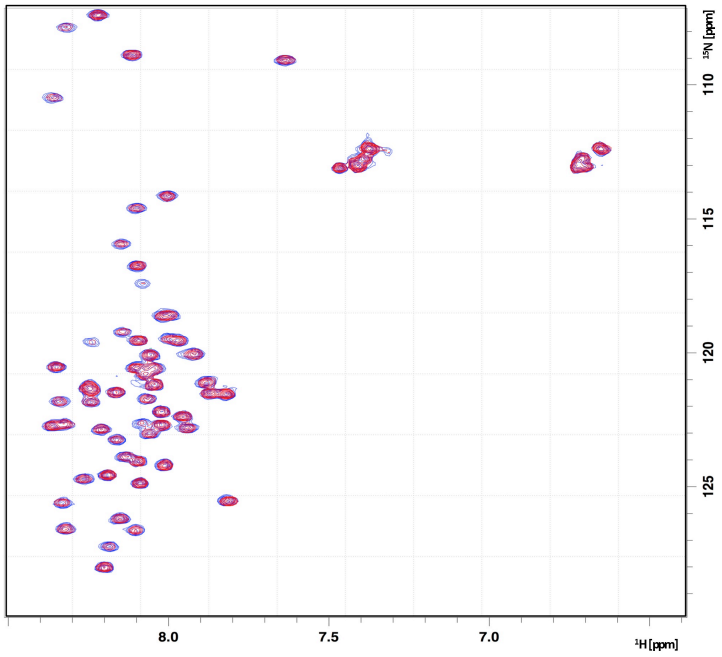



FIGURE 2.18: **NMR of $^{15}$N-PatE′ + PatG-DUF$_{di}$** $^{1}$H-$^{15}$N-HSQC of $^{15}$N-PatE′ before (blue) and after (red) addition of 2.0 equivalents of PatG-DUF$_{di.}$

.

The spectra of pure $^{15}$N PatE′ in the absence (blue) and presence of 2.0 equivalents of PatG-DUF$_{di.}$ (red) have been overlaid for clarity (Fig. 2.18), and clearly show an identical chemical shift pattern, showing that PatG-DUF$_{di.}$ does not bind the unmodified precursor peptide.

The NMR experiments were repeated using $^{15}$N PatE′ that had ben processed with PatD, introducing four heterocycles (two thiazolines and two oxazolines) within the context of the core peptide (Fig. 2.19). As before, the addition of up to 2.0 equivalents of PatG-DUF$_{di.}$ to the modified PatE′ resulted in an identical NMR spectrum to that of pure heterocyclised PatE′ (Fig. 2.20), indicating that the heterocycle-containing peptide does not bind PatG-DUF$_{di.}$ either. The known disorder of the leader peptide in PatE′ [51] would suggest that blocking of PatG-DUF$_{di.}$ binding by the leader peptide is unlikely. At present, we are unable to produce labelled cleaved precursor or macrocycle in sufficient quantities and purity for NMR experiments.

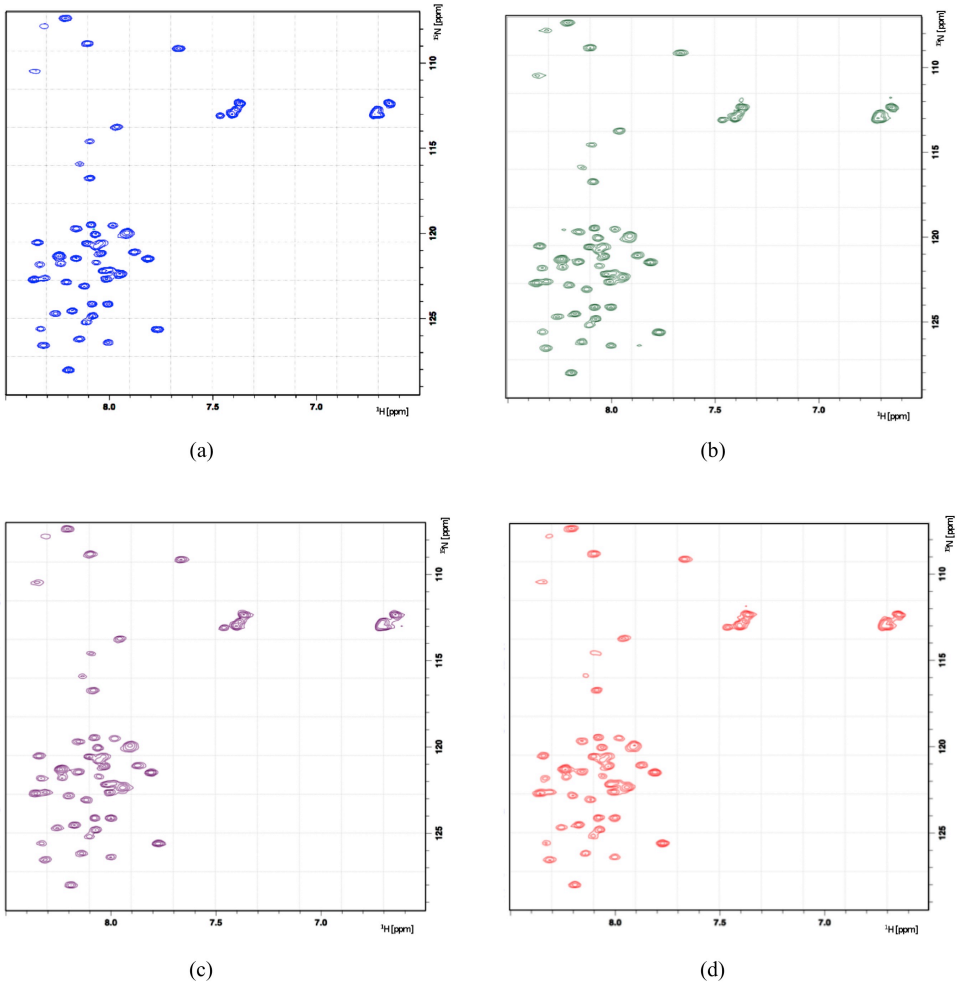



FIGURE 2.19: **NMR of heterocyclised $^{15}$N-PatE′ + PatG-DUF$_{di}$** $^{1}$H-$^{15}$N-HSQC of heterocyclised $^{15}$N-PatE′ (blue), with 0.5 (green), 1.0 (purple) and 2.0 (red) equivalents of PatG-DUF$_{di.}$.

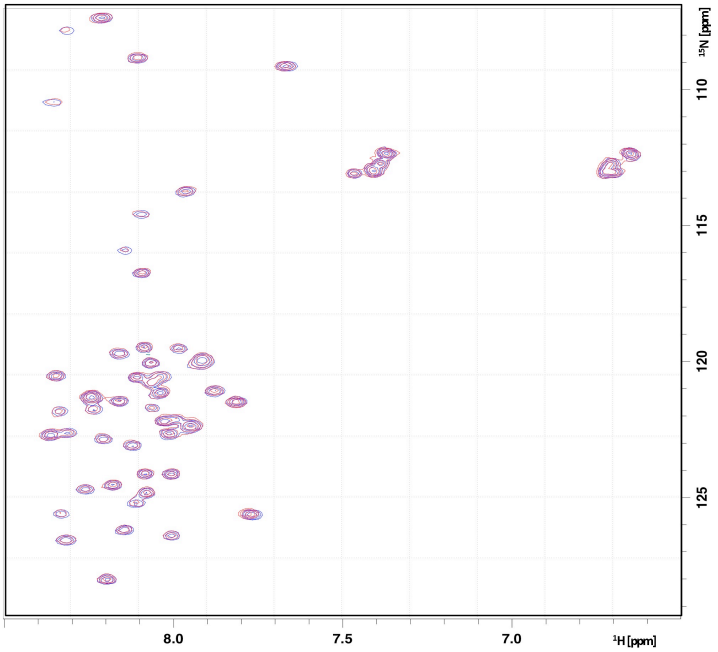



FIGURE 2.20: **NMR of heterocyclised $^{15}$N-PatE′ + PatG-DUF$_{di}$ overlaid** Heterocyclised $^{1}$H-$^{15}$N-HSQC of $^{15}$N-PatE′ before (blue) and after (red) addition of 2.0 equivalents of PatG-DUF$_{di.}$.

### 2.3.5 Investigating the Effect of PatG-DUF on Macrocyclisation

To investigate what effect, if any, the DUF has on peptide macrocyclisation, a new construct encompassing both the macrocyclisation and DUF domains of PatG, (hereby referred to as PatGmacDUF) was cloned into the pEHISTEV plasmid by Dr Jesko Koehnke and expressed and purified as described (section 2.2.10). The 74 kDa PatGmacDUF eluted from the gel filtration column as a single peak at approximately 80 mL, corresponding to a monomer (Fig. 2.21), yielding 1.5 mg L$^{-1}$ culture of pure protein, as assessed by SDS-PAGE (Fig. 2.21).

The rate of macrocyclisation was monitored using HPLC to follow the consumption of the IT$^{MeOx}$AC$^{ThH}$IT$^{MeOx}$AC$^{ThH}$AYDGE substrate peptide by both PatGmac and PatGmacDUF at 0, 2, 4, 6, 8, 10, 12, 24, 48 and 72 h time intervals. Start and end-points of the macrocyclisation reaction were established as 100 and 0 % of substrate respectively. No linear product was observed, so it can be assumed that complete consumption of substrate corresponds with complete product formation. Very little deviation between

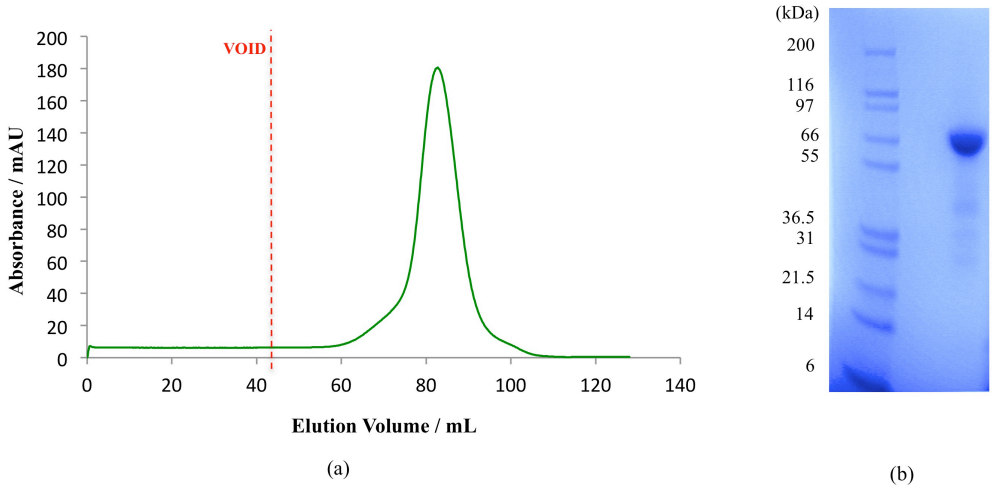



FIGURE 2.21: **Purification of PatGmacDUF** (a) Superdex S200 gel filtration chromatogram of PatGmacDUF. (b) SDS-PAGE analysis of PatGmacDUF compared with Mark12$^{TM}$ molecular weight marker. Identity was confirmed by MS.

the rates of macrocyclisation by PatGmac and PatGmacDUF was observed, and by 24 h both constructs have processed 90 % of the starting material, and both reactions were essentially complete (<1 and <5 % for PatGmac and PatGmacDUF respectively) by 48 h (Fig. 2.22).

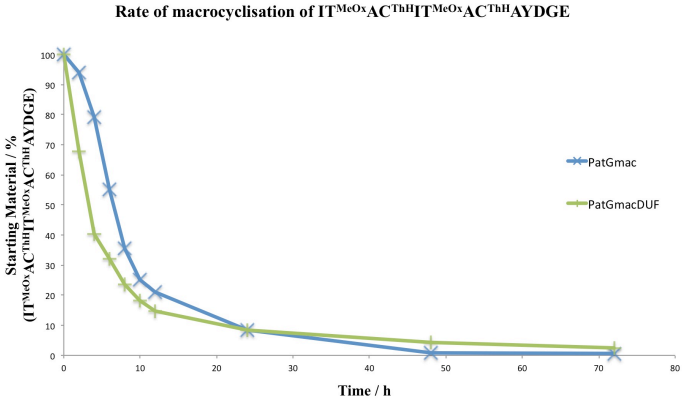



FIGURE 2.22: **Rate of macrocyclisation by PatGmac and PatGmacDUF** Consumption of IT$^{MeOx}$AC$^{ThH}$IT$^{MeOx}$AC$^{ThH}$AYDGE substrate during the macrocyclisation reaction over time by PatGmac (blue) and PatGmacDUF (green).

Any increase in the initial rate of macrocyclisation (0 - 12 h) observed upon the inclusion of PatG-DUF compared with PatGmac in isolation (although likely not significant) is offset by a much lower soluble protein yield following purification of PatGmacDUF compared with PatGmac (1.5 mg L$^{-1}$ and >200 mg L$^{-1}$ respectively)[43]. The increase

in activity is not sufficient to justify the use of PatGmacDUF as an alternative to the original PatGmac construct for biotechnological application.

## 2.4 Conclusions

The patellamide proteins PatA and PatG contain C-terminal domains of unknown function, which are 56 % identical and are conserved in related cyanobactin biosynthetic pathways. X-ray crystallography revealed the structure of PatG-DUF$_{sp.}$ to be a novel fold, and although a dimer is observed in the crystal, the residues involved in $Zn^{2+}$ coordination (which is required for dimerisation) are not well conserved in PatA-DUF or in homologues from other pathways. Consequently, it is difficult to speculate as to the functional importance of the dimer. We have been unable to detect any binding of PatG-DUF$_{di.}$ to linear substrates (simple peptides or peptides with heterocycles), indicating that these are unlikely to be substrates or ligands of the DUFs. The pitfall in using one technique to assess binding was shown by the reproducible, but functionally meaningless ITC binding curves that we observed. Further study is required with macrocyclic peptide before we can entirely rule out any binding between DUFs and patellamide substrates. It remains possible that the dimerisation of the domain indicates an interaction between DUFs, potentially leading to homodimers of PatA/PatG or even heterodimers of PatA and PatG. To investigate whether the DUFs affect the activity of their associated domains a construct containing both the macrocyclisation and the C-terminal DUF of PatG was designed. No significant enhancement of activity was observed for PatGmacDUF over PatGmac *in vitro* to offset the far lower soluble yield of PatGmacDUF compared with PatGmac, limiting its potential for biotechnological application. If the DUFs have no enzymatic role, which our lack of binding data would suggest, coupled with the observation that the protease domains of PatA and PatG are active in isolation, which is not significantly altered by the presence of the DUFs, it is interesting as to why they are conserved throughout cyanobactin biosynthesis. However, importantly, it also suggests that the DUFs can be excluded from *in vitro* enzyme-based synthesis systems. This simplifies the design of a biotechnological 'toolkit' for bioactive cyclic and linear peptides.

## 2.5 Future Work

Although the work in this chapter demonstrates that the DUFs are not required for the synthesis of cyanobactins *in vitro* a role for the DUFs is yet to be determined. Efforts to identify a role for DUFs should begin by monitoring for a possible binding interaction between PatG-DUF and short, leaderless peptides and macrocyclic peptides. Additionally, it remains possible that the DUFs are scaffold for multi-protein complexes, and this is to be investigated.

# Chapter 3

# Structure-guided Characterisation and Engineering of the Heterocyclase

## 3.1 Introduction

The incorporation of heterocycles within a peptide backbone, significantly alters the peptide's physiochemical properties, giving rise to more medicinally favourable characteristics, including increased membrane permeability, and reduced conformational flexibility, which can both improve oral bioavailability and target affinity[99]. Therefore, enzymes offering a straightforward and site-selective installation of heterocycles into a peptide backbone are a desirable and applicable biotechnology. The cyanobactin heterocyclase enzymes (D enzymes) catalyse the ATP-dependant cyclodehydration of cysteine, and in some cases serine/threonine residues, forming thiazoline and oxazoline heterocycles within the context of the cyanobactin core peptide. These enzymes show high substrate tolerance, selectively modifying these residues, largely irrespective of their neighbouring residues, suggesting substrate recognition is spatially distinct from catalysis. The importance of the highly conserved leader peptide in substrate processing is well documented, and is thought to be involved in recognition[41,51,52]. However at the time of writing the molecular rational underpinning this recognition is unknown. Additionally, at present, the precise utilisation of ATP during turnover is contentious,

with evidence supporting its involvement in substrate activation via both the $\alpha$ and $\gamma$-phosphate[50,51].

In this chapter, we make use of X-ray crystal structures of the cyanobactin heterocyclase LynD (highly homologous to PatD) from the aestuaramide pathway (*Lyngbya sp.*) in complex with various substrate peptides and nucleotides (determined by Dr Jesko Koehnke) to provide insight into substrate recognition and the mechanism of heterocyclisation. Analysis of the crystal structures reveal the orientation of the nucleotide bound at the active site is not consistent with an adenylation mechanism[51]. Consequently the utilisation of ATP as a cofactor is looked at in more detail. The complex structures also provide direct evidence for the role of the leader peptide in substrate recognition, rationalising the observed *cis* and *trans* activation of the heterocyclase[41] by the leader peptide, and the sequence of the 'minimal leader' required for complete processing[51,52]. Using these data, a novel LynD has been engineered, capable of affecting a short peptide lacking the leader peptide, with wild-type-like activity.

## 3.2 Materials and Methods

### 3.2.1 Expression and Purification of LynD, AcLynD and LynD Mutants

Codon-optimised full-length LynD (*Lyngbya* sp. PCC-8106) with an N-terminal $His_6$ tag and TEV protease site was purchased from DNA2.0 in the pJexpress 411 plasmid. Mutagenesis was performed by Dr Andrew Bent using established protocols[80]. For biochemical characterisation all proteins were expressed and purified as described (section 2.2.6).

### 3.2.2 Expression and Purification of PatE′ and PatE′ mutants

PatE′ and PatE′ mutants were expressed and purified as described (section 2.2.5). Mutagenesis was performed by Dr Andrew Bent using established protocols[80].

### 3.2.3 Synthesis of a Detectable Full-length Leader Peptide

To generate full-length leader peptide, the PatE′ sequence was mutated at four positions (D2Y, K3E, K4E and R16E) using standard protocols[80]. The resulting quadruple mutant (PatE′4) full amino acid sequence is as follows: MYEENILPQQGQPVIELT-AGQLSSQLAELSEEALGDAGLEASKITACITFCAYDGELEHHHHHH. PatE′4 was expressed and purified as described for PatE′ (2.2.5) and subjected to heterocyclisation tests to ensure normal processing. The mutations K3E, K4E and R16E enable full-length leader to be retained following digestion of PatE′4 with trypsin (Sigma-aldrich) and the D2Y mutation allows the leader peptide to be visualised during purification from the trypsin reaction and quantified. PatE′4 was digested with 1/100 trypsin at 37 °C, 300 rpm for 3 h and subsequently applied to a Ni-Sepharose Fast Flow column (GE Healthcare) equilibrated in gel filtration buffer. The flow-through was collected and was confirmed to contain full-length PatE′4 leader peptide by MALDI-TOF-MS. The pure sample was concentrated to 1 mM.

### 3.2.4 Heterocyclisation Reactions

Standard heterocyclisation reaction conditions are as follows: 100 μM PatE′ (and variants) were incubated with 5 μM enzyme, 5 mM ATP, 5 mM $MgCl_2$ in gel filtration buffer at 37 °C for 16 h. Reactions set up for analysis of ATP hydrolysis by $^{31}P$ NMR were prepared using 220 μM ATP and 1 mM DTT in place of 1 mM TCEP (from gel filtration buffer); the reaction was incubated at 37 °C for 2 h. To investigate the activity of LynD and variants towards leaderless substrates, 5 μM enzyme was incubated with 100 μM synthetic ITACITFCAYDG in the presence and absence of 5.5 μM synthetic LAELSEEAL (minimal leader) or full-length PatE′4 leader peptide. The reactions were otherwise performed under standard conditions. Synthetic peptides were purchases from Peptide Protein Research Ltd. Samples were analysed either by LC-ESI-MS or MALDI-TOF-MS.

### 3.2.5 Comparative Time-Course Experiments

To assess the relative rates of heterocyclisation under various conditions, reactions were monitored at regular intervals using MALDI-TOF-MS. In each case reactions of PatE′

or ITACITFCAYDG were prepared as described (3.2.4) and incubated at 37 °C prior to the addition of LynD (wild-type, mutants and fused), allowing a 0 time point to be recorded. Time points for each reaction were as follows: for PatE′ with LynD (standard conditions) and ITACITFCAYDG with LynD fusion, the reaction was monitored after 1, 5, 10, 15, 20, 30, 45, 60, 90, 120 and 180 min. For ITACITFCAYDG with LynD in the absence and presence of either the minimal leader LAELSEEAL peptide or full-length PatE′4 leader peptide, the reaction was monitored after 15, 30, 60, 120, 240 and 1440 min. For PatE′ with wild-type LynD and $LynD^{K409A}$ with 500 μM ATP the reaction was monitored after 10, 30, 60, 120 and 360 min. Reactions were set up in triplicate and each sample was analysed by MALDI-TOF-MS in triplicate (nine spectra recorded per time point). For each time point the total ion count for each species (0, 1 and 2 heterocycles) was recorded and averaged, and the percentage of each species was calculated. Plotting the percentage of species containing 1 and 2 heterocycles against time allowed a visual comparison for the rate of the reaction relative to the wild-type reaction (PatE′ + LynD).

### 3.2.6 ITC Data Collection and Analysis

ITC experiments were performed using a VP-ITC instrument (MicroCal) in gel filtration buffer supplemented with 5 mM $MgCl_2$. LynD and PatE′ (and their respective mutants) were dialysed into freshly prepared buffer overnight at 4 °C to ensure the buffers for the experiment were matched. For substrate binding experiments a cell solution of 10 μM LynD (and mutants) and a syringe solution of 150 μM PatE′ (and mutants) were prepared by diluting protein and peptide with dialysis buffer. For nucleotide binding experiments, the cell concentration was increased to 20 μM, and the ATP/AMP concentration in the syringe solution was 300 μM. Substrate and ATP experiments were performed at 20 °C and AMP experiments were performed at 25 °C. Cell and syringe solutions were degassed for 15 min at 18 and 23 °C respectively. The titration method was as follows: one injection of 2 μL followed by injections of 5 μL at 0.5 μL $min^{-1}$, with a delay of 4 min between injections. A stirring speed of 307 rpm was used throughout. Raw data were processed using MicroCal *Origin* software, the baseline was adjusted and the integration limits were set manually. Data were fitted to the one-site model (*Origin*), setting the stoichiometry to 1.

### 3.2.7 $^{31}$P NMR Experiments

$^{31}$P NMR spectra were acquired using 1,024 scans on a Bruker AVANCE III 500 MHz spectrometer equipped with a room temperature BBFO+ probe. Composite pulse sequence 'waltz-64' was used for $^{1}$H power-gates broadband decoupling.

### 3.2.8 Selective $^{18}$O-PatE′

PatE′ was selectively $^{18}$O labelled at the carbonyl of the residues preceding the cysteine residues based on a protocol described in the study of BalhD[50]. PatE′ was heterocyclised as described (section 3.2.4) and the reaction mixture was purified using size-exclusion chromatography (Highload 16/60 Superdex 75, GE Healthcare). Pure heterocyclised-PatE′ was concentrated to 0.5 mL and dried to a powder using a Thermo Savent Speed Vac. To accomplish selective $^{18}$O labeling, the dried sample was resuspended in $H_2{}^{18}O$ supplemented with 0.1 % formic acid to hydrolyse the heterocycles, and incubated at 20 °C. The ring opening was followed using MALDI-TOF-MS. Once complete, selectively labelled $^{18}$O-PatE′ was dialysed into gel filtration buffer (containing 1 mM DTT in place of 1 mM TCEP). To ensure all traces of acid and $H_2{}^{18}O$ were removed, the sample was dialysed 3 x 1:1000 (2 x 2 h followed by overnight). The dialysed sample was recovered, diluted to 7.5 mL in urea lysis buffer supplemented with 10 mM DTT and passed through a Highload 16/60 Superdex S75 size-exclusion column (GE healthcare) pre-equilibrated in gel filtration buffer containing 1 mM DTT in place of 1 mM TCEP to yield pure $^{18}$O-PatE′, as confirmed by MALDI-TOF-MS.

### 3.2.9 LC-ESI-MS of $^{18}$O-PatE′ pyrophosphate

Purification of pyrophosphate was achieved by passing the reaction mixture through a protein concentrator with a 5 kDa cutoff. The flow-through was applied to CaptoQ resin (GE Healthcare) and washed with 15 column volumes of $dH_2O$. Pyrophosphate was eluted with 1M sodium acetate, pH 7.0. Accurate mass measurements were carried out on an ABSciex 5600 mass spectrometer with a nanospray source. Samples were diluted 50 % with acetonitrile and infused at 1 μL min$^{-1}$ using the built-in syringe pump. Negative ionization spectra were collected from 60-400 m/z in MS mode. Mass/charge (m/z) ratios of 176.9 and 178.9 were isolated in Q1 and fragmented with collision energy

of 30 V in Q2, and the fragments were measured from 50-200 m/z on the TOF analyser. The TOF was externally calibrated with a mixture of periodate (190.8847 m/z) and sulfanilic acid (172.0074 m/z).

### 3.2.10 Preparation of Test Substrates for AcLynD

PatE′ variants with seven (PatE′-IACIMAC) and nine (PatE′-IITACIMAC) core peptide residues, and one with three cysteine residues (PatE′-ICACITFC) in the core peptide were cloned, expressed and purified as described for PatE′ (section 2.2.5). Each peptide was digested with 1/100 trypsin at 37 °C, 300 rpm for 3 h and subsequently purified using size-exclusion chromatography (Highload 16/60 Superdex S30, GE Healthcare) in gel filtration buffer containing an additional 1 mM TCEP (2 mM total).

## 3.3 Results and Discussion

The cyanobactin heterocyclase LynD from the aestuaramide pathway (*Lyngbya sp.*) is a close homologue of the patellamide heterocyclase PatD. To date, only products from the aestuaramide pathway containing thiazole heterocycles have been identified[100], suggesting LynD would display TruD-like activity, processing only cysteine residues within a core peptide (compared to PatD, which additionally processes serine and threonine residues)[40]. However, since the core peptides from which the aesuaramides are derived, lack serine and threonine residues, this hypothesis needed to be tested. While the core peptide sequence of the aesuramide and patellamide precursor peptides are varied, the conserved recognition elements and leader peptide from the two pathways are highly similar, and so it is expected that PatE′ would be a substrate of LynD (Fig. 1.2).

Incubation of LynD with PatE′ in the presence of ATP/$Mg^{2+}$ overnight (to ensure completion), resulted in a species with a mass loss of 36 Da as identified by LC-ESI-MS, corresponding with two heterocyclisation events (Fig. 3.1). Iodoacetamide (IAA), a molecule that forms adducts with free cysteine was added to the heterocyclised product, and no mass increase was detected by LC-ESI-MS, confirming the loss of 36 Da corresponded with the modification of the two cysteine residues within the core peptide (Fig. 3.1).

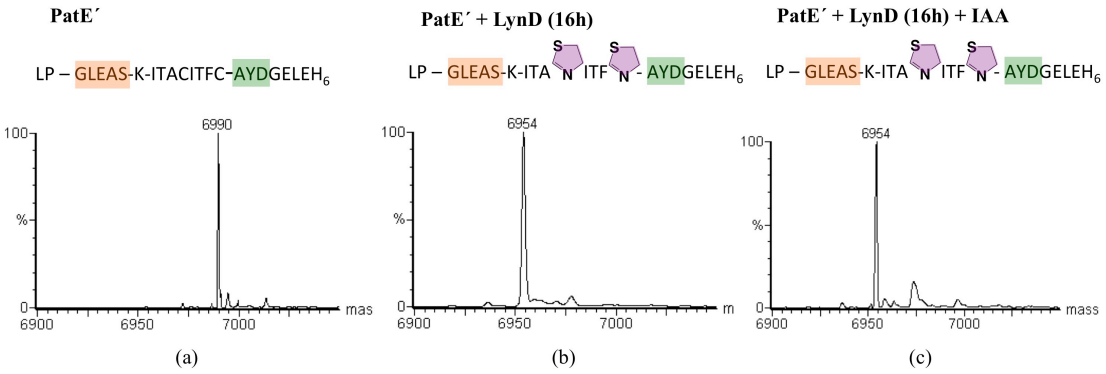
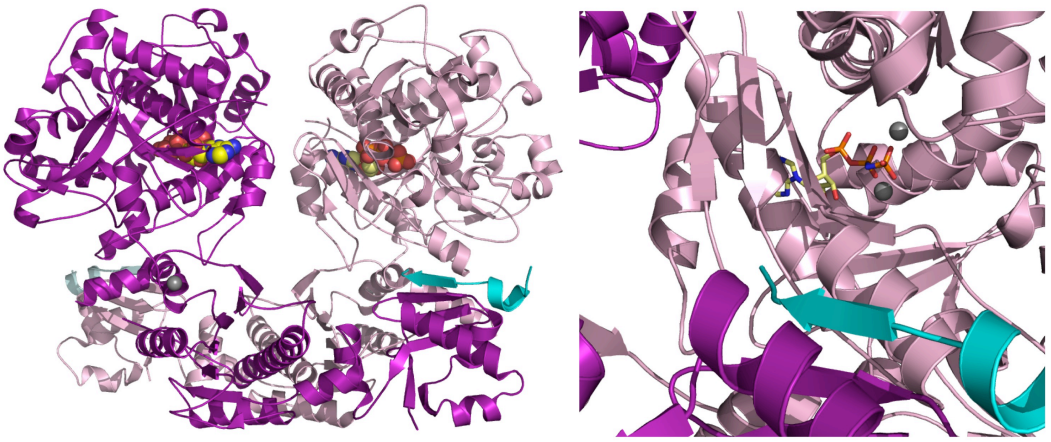



FIGURE 3.1: **PatE′ + LynD LC-ESI-MS** of (a) PatE′, (b) PatE′ + LynD and (c) PatE′ + LynD + Iodoacetamide (IAA). The loss of 36 Da (two water molecules) coupled with the lack of mass increase following incubation with iodoacetamide, consistent with the two core peptide cysteine residues have undergone cyclodehydration, forming two thiazoline heterocycles.

### 3.3.1 Structure of LynD Complexes

X-ray crystallography of LynD in complex with variants of substrate peptide PatE′ and various nucleotides were pursued by Mr Hannes Ludewig and Dr Jesko Koehnke. Three complex structures were solved by Dr Jesko Koehnke: LynD-AMP-PatE′, 2.86 Å, (PDB: 4V1U); LynD-ATP-PatE′$^{C51A}$, 2.14 Å, (PDB: 4V1T); LynD-β,γ-imido-ATP-PatE′$^{C51A}$ (LynD-AMP-PNP-PatE′$^{C51A}$), 3.01 Å, (PDB: 4V1V). All three complexes contained two monomers in the asymmetric unit, with each monomer composed of three domains, which form an antiparallel dimer through head-to-tail association of domains 1 and 2, as seen for TruD (PDB: 4BS9)[51]. The LynD complex structure is overall tighter than that seen for the apo TruD structure; particularly domains 3 from each monomer are much closer together in the complex structure (Fig. 3.2). All three complex structures are highly similar, with low Cα r.m.s.d. values ($\leq$ 1.06 Å) between the three structures. The following discussion refers to the highest-resolution LynD-ATP-PatE′$^{C51A}$ structure, except when specified. In this model LynD residues 6-143, 151-229, 240-336 and 342-775 in chain A and residues 4-226, 240-336 and 343-775 in chain B are ordered. The missing residues are presumed disordered. A $Zn^{2+}$ is coordinated in domain 2 in the same manner as TruD[51]. The structure allows us to locate both the active site and substrate recognition site. The nucleotide at the active site is in domain 3, whereas the leader peptide is bound at the interface between domain 1 from one monomer and domain 3 from another (Fig. 3.2). Simple distance constraints dictate that processing

of the core peptide must occur in the same domain 3 that binds the leader peptide. The nucleotide in this domain is approximately 26 Å away from the closest ordered residue of the leader peptide, sufficiently close to position the subsequent core peptide near to this nucleotide. Conversely the shortest accessible route from the leader peptide to the alternate nucleotide gives a total distance of over 70 Å; the precursor peptide is not long enough to reach this nucleotide. Therefore the dimer is a functional requirement (Fig. 3.2).



(a) (b)

FIGURE 3.2: **X-ray crystal structures of LynD complexes** (a) Cartoon representation of LynD-ATP-PatE′$^{C51A}$ complex showing dimeric assembly. LynD (purple) is comprised of the same 3 domains as seen for TruD [51]. A structural $Zn^{2+}$ ion (grey sphere) is coordinated in domain 2 and the nucleotide (yellow spheres) is bound in domain 3. The ordered residues of PatE′$^{C51A}$ (cyan cartoon) is bound between domain 1 of one LynD monomer, and domain 3 of another. The second monomer is shown using the same colour scheme as the other, but using faded colours. (b) Cartoon representation of LynD-AMP-PNP-PatE′$^{C51A}$ (LynD monomers are shown in purple and light pink). The nucleotide nearest to the last ordered residue of PatE′$^{C51A}$ (therefore the catalytic nucleotide for that peptide), is located in the same domain 3 which makes contacts with the leader peptide, indicating the dimer is a functional requirement. A clear path from the last ordered residue of PatE′$^{C51A}$ to AMP-PNP can be seen. AMP-PNP is shown as yellow stick. Magnesium ions in the active shown as grey spheres.

### 3.3.2 Nucleotide Binding Site

The nucleotide binding site, and thus the active site of LynD is located within domain 3 (residues 316-775). This is inconsistent with the hypothesis that the cyanobactin heterocyclase enzymes are adenylating enzymes, which was in part proposed due

to structural similarity between the cyanobactin heterocyclase TruD, and MccB, an adenylase required for the biosynthesis of the antibiotic microcin C7, which would require ATP to be bound in domain 2 (although the authors did note that the ATP binding site of MccB is not fully conserved in TruD)[51]. Domain 3 of LynD is structurally homologous to the *E. coli* YcaO domain[101] (C$\alpha$ r.m.s.d. of 2.4 Å over 282 residues) and binds the nucleotide in the same location. The adenosine ring is bound through cation-$\pi$ stacking interactions with R344 on one face, and makes van der Waals interactions with S419 and T351 on the other. Additionally it forms hydrogen bonds through the N7 and N1 atoms with residues Q415 and N536 respectively (Fig. 3.3). The lack of extensive hydrogen bonding around the ring is consistent with the ability of the D enzymes to use other nucleotide triphosphates[51]. The O2′ and the O3′ atoms of the ribose hydrogen bond with the main chain of A534 and the side chain of E426 respectively. The $\alpha$-phosphate makes a salt bridge with R636 and hydrogen bonds with Q544 and S419. In the LynD-ATP-PatE′$^{C51A}$ structure, a discontinuation of the nucleotide electron density caused us to reassign the complex more accurately as having bound ADP and phosphate ($P_i$) (Fig. 3.3). It is presumed the $P_i$ mimics the $\gamma$-phosphate of ATP. It is unclear as to whether the hydrolysis of ATP to ADP and $P_i$ in the crystal structure is mechanistically relevant, but if so it would be consistent with substrate activation via a phosphorylation mechanism[49,50]. The precise positions of the $\beta$ and $\gamma$-phosphates vary slightly between the ADP-$P_i$ and the AMP-PNP structures, but their similarities are informative. In both complexes a $Mg^{2+}$ ion, coordinated by E640, binds the $\beta$ and $\gamma$-phosphates and a second $Mg^{2+}$ ion, coordinated by E548, binds the $\beta$-phosphate. In the ADP-$P_i$ structure the second $Mg^{2+}$ bridges the $\beta$ and $\gamma$-phosphates (Fig. 3.3). This structure also contains a third $Mg^{2+}$ ion, coordinated by E423, which also bridges the $\beta$ and $\gamma$-phosphates. In addition to the metal ions, the $\beta$-phosphate makes salt bridges with R636 and K409, and the $\gamma$-phosphate forms a salt bridge with R552 and R427 (Fig. 3.3).

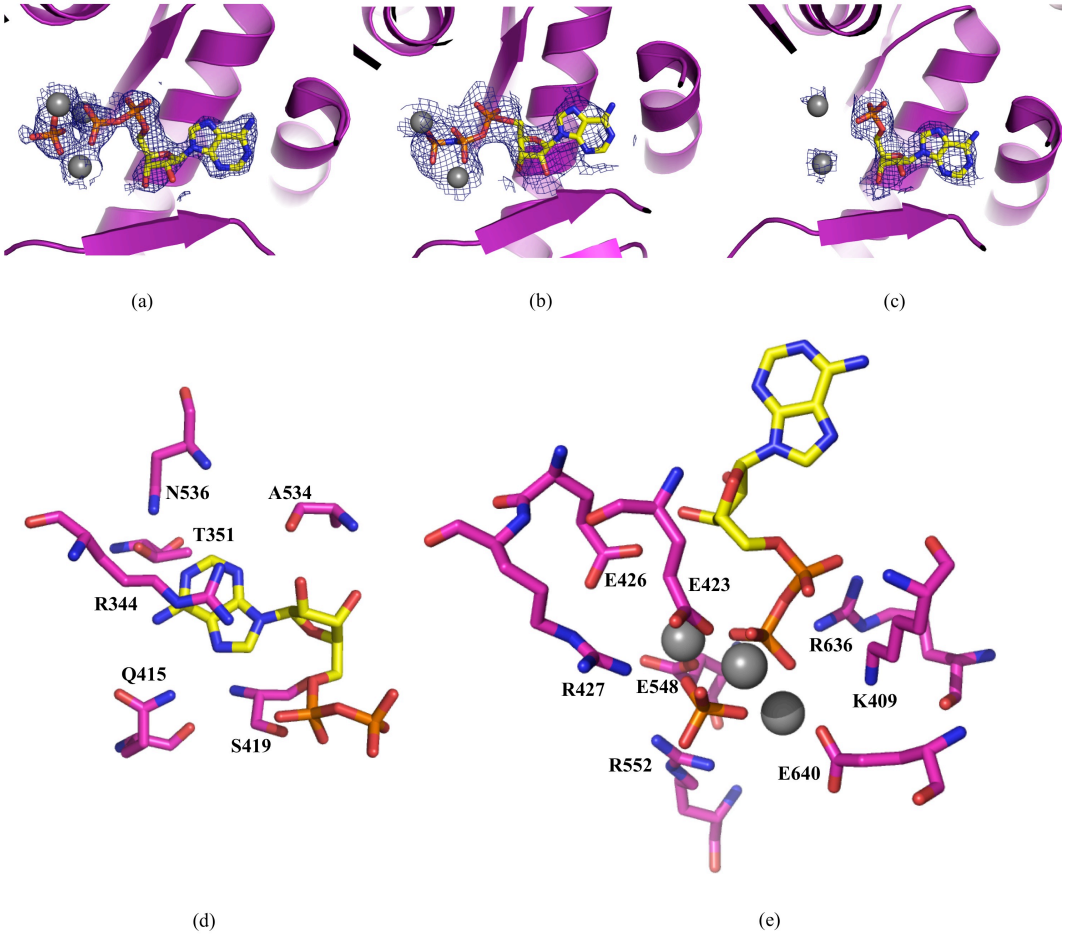
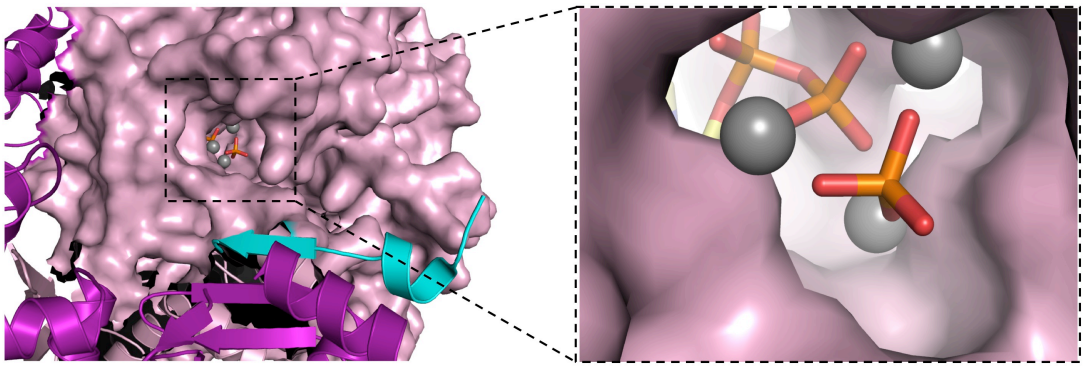



FIGURE 3.3: **LynD nucleotide binding site** (a) LynD-ADP-$P_i$-PatE′$^{C51A}$, (b) LynD-AMP-PNP-PatE′$^{C51}$ and (c) LynD-AMP-PatE′ complex structures. Difference electron density ($F_0$-$F_c$) contoured at 3 σ with phases calculated from a model that was refined without nucleotide or metal ions present, shown as a blue isomesh. (d) Residues involved in binding of adenosine and ribose moeity of ADP shown as sticks. (e) Residues involved phosphate and metal ion coordination shown as sticks. Colouring for all images are as follows: Nucleotide atoms are coloured yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorous. Protein atoms follow the same colour scheme, except carbons are purple. Metal ions are shown as grey spheres.

### 3.3.3 Orientation of Nucleotide Binding

Heterocyclisation of cysteine, serine and threonine proceeds via attack on the nucleotide triphosphate by a hemiorthoamide intermediate. However these complex structures reveal that the orientation of the nucleotide within the active site, relative to the incoming substrate is such that the α-phosphate is occluded from attack by the hemiorthoamide (Fig. 3.4), thus precluding an adenylation mechanism[51].



(a) (b)

FIGURE 3.4: **Occlusion of α-phosphate** The α-phosphate is occluded from attack of the reacting hemiorthoamide intermediate. (a) Incoming precursor peptide threads through to the active site, through the channel as shown, approaching the γ-phosphate. (b) Close-up view of the nucleotide in the active from the perspective of the incoming peptide. From this angle shielding of the α-phosphate clearly precludes an adenylation mechanism. One monomer of LynD is shown as a purple cartoon. Domain 3 of the other monomer is represented with a pink surface. The precursor peptide is shown as a cyan cartoon. Nucleotide phosphates are shown as sticks coloured orange for phosphorous and red for oxygen. Metal ions are shown as grey spheres.

To ensure that the observed orientation of the nucleotide is mechanistically relevant and to eliminate any possibility that it is an artefact of crystallisation, residues identified as nucleotide binding were mutated, and the mutants were tested for their ability to bind ATP or AMP and process substrate PatE′. Binding experiments were performed using ITC and can be seen in Fig. 3.5. The activity of each mutant was monitored using MALDI-TOF-MS. A summary of the binding and activity data can be seen in Table 3.1.

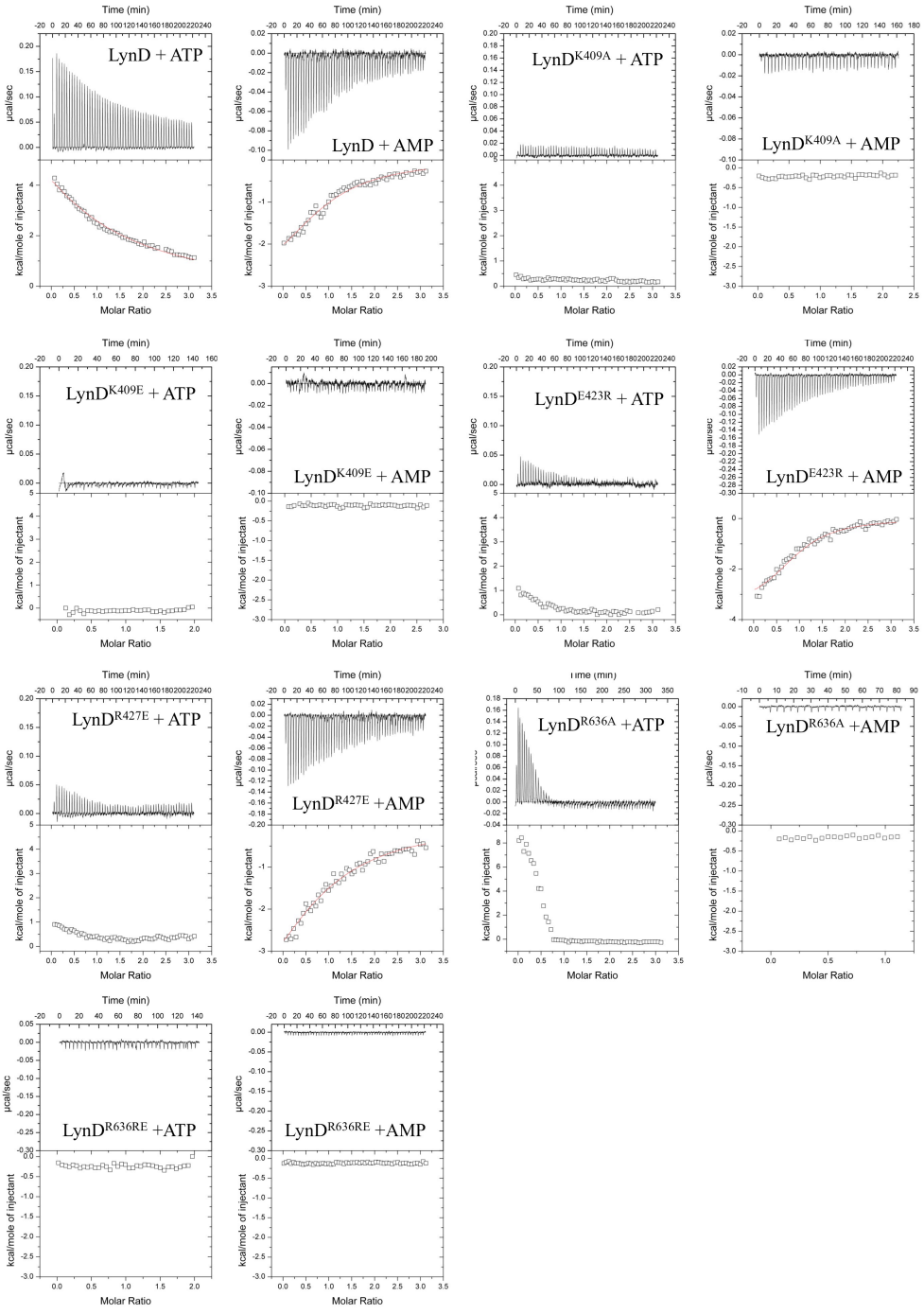



FIGURE 3.5: **Nucleotide binding** ITC data for ATP and AMP titrated into LynD and LynD mutant solutions. The top panel shows raw data representing heat evolved in response to injections, the bottom panel shows the integrated heats of injections (□) and the best fit (-) to the one-site model (*Origin*).

TABLE 3.1: **Nucleotide binding and processing of LynD mutants**.

| | Nucleotide $K_D$ (μM) | | |
|---|---|---|---|
| **Enzyme** | **ATP** | **AMP** | **No. of heterocycles (PatE′)** |
| LynD | 50.25 | 14.41 | 2 |
| $LynD^{K409A}$ | No binding | No binding | 2 |
| $LynD^{K409E}$ | No binding | No Binding | 1 |
| $LynD^{E423R}$ | No binding | 5.65 | 0 |
| $LynD^{R427E}$ | No binding | 23.04 | 1 and 2 |
| $LynD^{R636E}$ | No binding | No binding | 0, 1 and 2 |
| $LynD^{R636A}$ | N/A | No binding | 0, 1 and 2 |

Both the K409A and K409E mutants completely abolished binding of ATP and AMP to LynD as determined by ITC experiments (Fig. 3.5, Table 3.1). Therefore the ability of each mutant to process PatE′ substrate was unexpected (Table 3.1). The K409A mutant resulted in complete substrate turnover, converting both cysteine residues to thiazolines, while the K409E mutant slowed processing, catalysing only one cyclodehydration under standard conditions (Table 3.1). These data imply that both the K409A and K409E mutants do bind ATP, but only very weakly, below the detection limits of ITC, and that our standard reaction conditions contain sufficient ATP (5 mM) to facilitate catalysis. Consequently the ability of the K409A mutant to process substrate PatE′ was investigated again, using a lower ATP concentration (500 μM), and the heterocyclisation reaction was monitored over time using MALDI-TOF-MS. By plotting the amount of product formed over time, as identified by MALDI-TOF-MS under these new reaction conditions, it becomes obvious that the K409A mutation impairs processing, forming both heterocycles more slowly compared with wild-type enzyme (Fig. 3.6).

The R636E and R636A mutations inhibit binding of both ATP and AMP to LynD (Fig. 3.5). Unfortunately the ITC experiment between $LynD^{R636A}$ and ATP reproducibly resulted in the complex ITC curve as shown (Fig. 3.5), and could not be easily interpreted, and at present we are unable to explain this result. Both mutants significantly affect substrate processing, resulting in a mixture of products containing

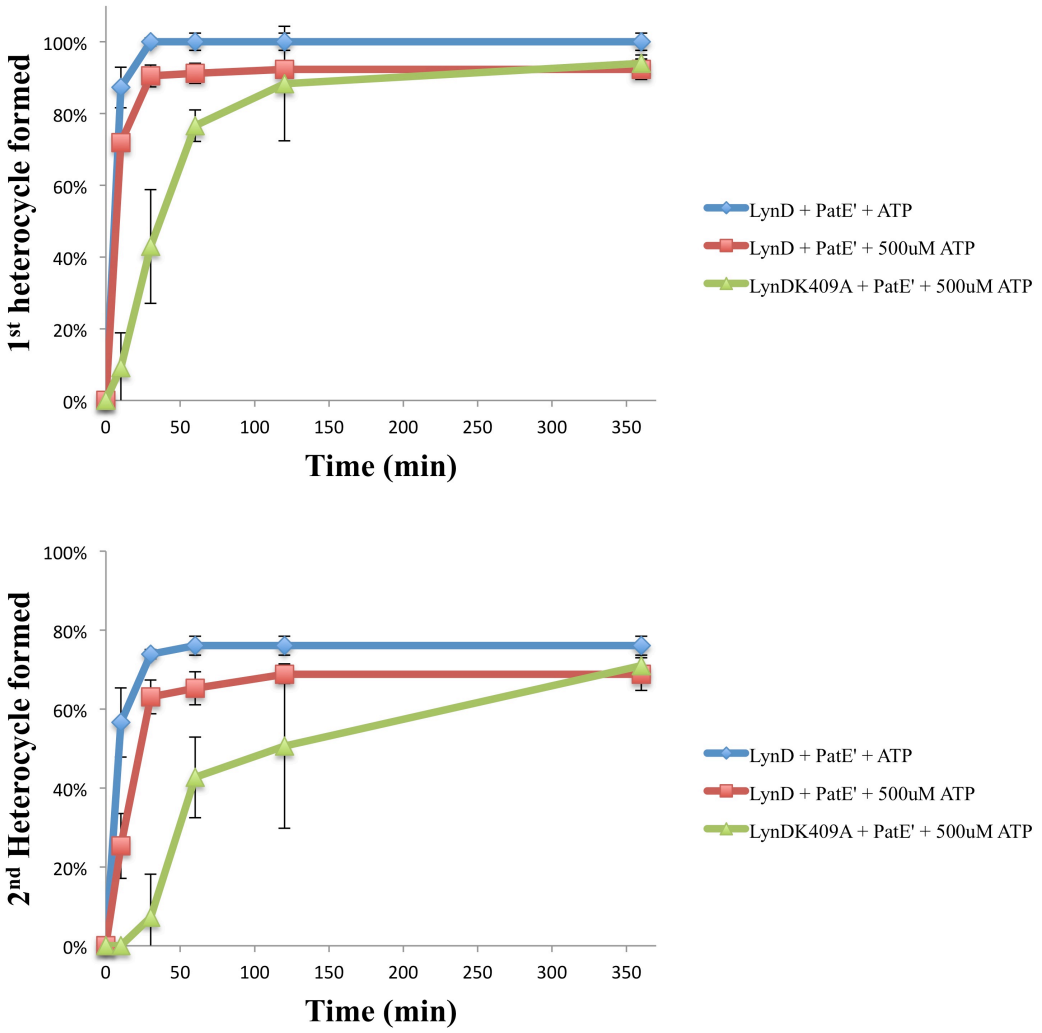



FIGURE 3.6: **LynD$^{K409E}$ heterocyclisation rate** Relative rates of the heterocyclisation reaction between PatE′ and LynD + 5 mM ATP (blue), PatE′ and LynD + 500 µM ATP (red) and PatE′ and LynD$^{K409E}$ + 500 µM (green) analysed by MALDI-TOF-MS. The top graph shows the time taken to form the 1st heterocycle and the bottom graphs shows the time taken to form the 2nd heterocycle. Each experiment was set up in triplicate and each measurement was repeated three times; thus each time-point is an average of nine measurements. Errors are plotted as +/- 1 s.d..

0, 1 and 2 heterocycles. These data support the idea that R636 coordinates the α-phosphate, as observed in the crystal structure.

The E423R and the R427E mutations also impair substrate processing, forming 0 heterocycles, and a mixture of 1 and 2 heterocycles respectively, under standard reaction conditions. Interestingly, while both mutations completely abolish ATP binding, AMP

still binds to each mutant (Fig. 3.5, Table 3.1). The E423R mutant binds AMP with higher affinity, and the R427E mutant shows only slightly weaker affinity than the wild-type enzyme (Table 3.1). The ability of the E423R and the R427E mutants to bind AMP, but not ATP, confirm the involvement of these residues in the coordination of β and γ-phosphates as predicted by the crystal structure, indicating the observed orientation of the nucleotide in the active site is correct, and not an artefact of crystallisation. Consequently these data, strongly suggest heterocyclisation does not proceed via the previously proposed adenylation mechanism[51].

### 3.3.4 Nucleotide Utilisation

Simply ruling out an adenylation mechanism for heterocyclisation does not by extension confirm the alternative hypothesis, that substrate is activated via a kinase mechanism, as proposed for the LAP heterocyclase BalhD[49,50]. An adenylation mechanism was not proposed solely because of structural homology between TruD and an adenylating enzyme, but due to the direct observation (by $^{31}$P NMR) that AMP and pyrophosphate ($PP_i$) and phosphate ($P_i$) are produced during turnover[51]. While the presence of AMP could be explained by the degradation of ADP following substrate phosphorylation by ATP; a kinase mechanism is not sufficient to explain the presence $PP_i$[51]. Additionally, in the TruD study, at no point was ADP detected[51]. It is unlikely the cyanobactin and the LAP classes of heterocyclase operate via different mechanisms, so we decided to look at the utilisation of ATP in the cyclodehydration reaction in more detail, in an attempt to resolve these mechanistic ambiguities.

Using $^{31}$P NMR, LynD, was observed to produce AMP, $PP_i$ and $P_i$ both in the absence of substrate (similarly to the *E. coli* YcaO domain[101]) and in the presence of substrate PatE′ (like TruD[51]). ADP was not detected (Fig. 3.7).

For TruD, the production of AMP, $PP_i$ and $P_i$ was correlated to the appearance of heterocycles, and not just background ATP hydrolysis[51]. It was assumed that the $P_i$ arose due to the degradation of $PP_i$ under experimental conditions and that AMP and $PP_i$ were the mechanistically relevant biproducts of cyclodehydration[51].

To investigate this further, we sought to follow on from an experiment described in the BalhD study, which utilised $^{18}$O at the carbonyl of the residues directly preceding the

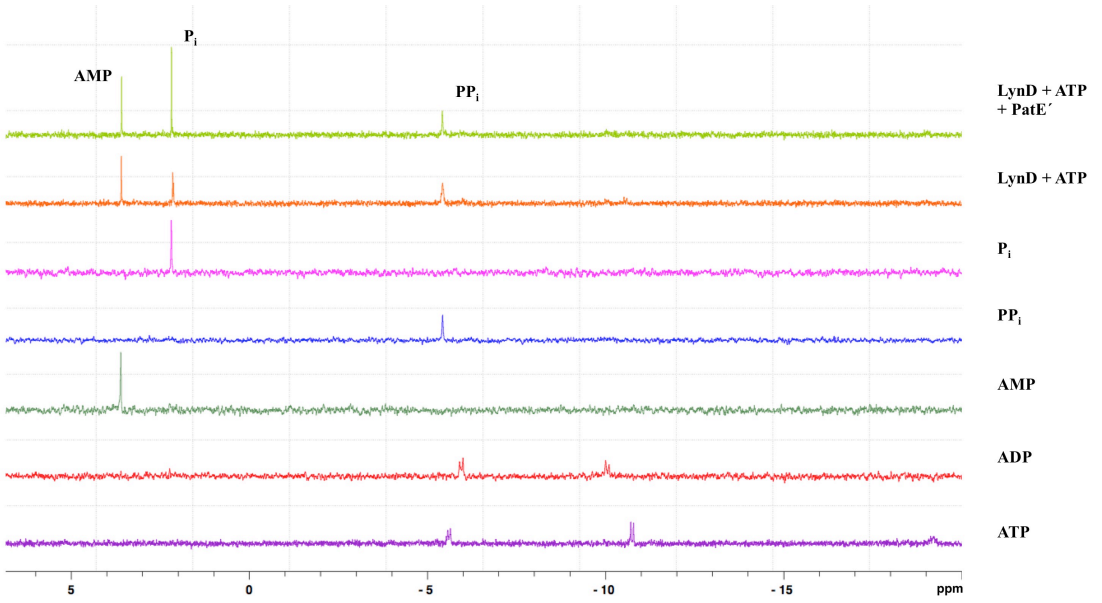
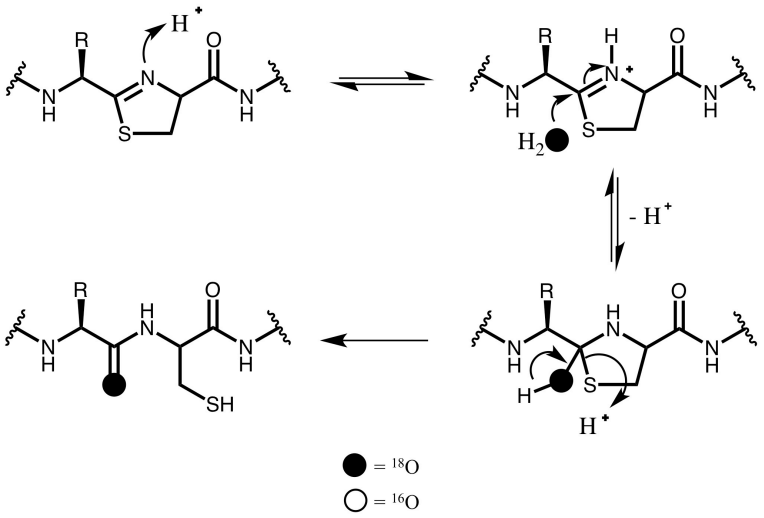



FIGURE 3.7: **$^{31}$P NMR spectra of LynD reactions** $^{31}$P NMR spectra of LynD + ATP + PatE′ reaction (green), LynD + ATP background reaction (orange), $P_i$ (pink), $PP_i$ (blue), AMP (teal), ADP (red) and ATP (purple). Comparison between LynD reactions and standards show LynD hydrolises ATP to AMP, $PP_i$ and $P_i$. No ADP is detected.

cyclisable cysteines, and determined the role of ATP in substrate activation due to the incorporation of the $^{18}$O in $P_i$, as identified by an isotope shift of the $P_i$ peak in a $^{31}$P NMR spectrum[50]. To generate such a selectively $^{18}$O-PatE′ variant, PatE′ was first heterocyclised as described, purified and lyophilised. The dried sample was re-solubilised in $H_2{}^{18}O$ supplemented with 0.1 % formic acid, and incubated at 20 °C to hydrolise the heterocycles and accomplish selective $^{18}$O labelling. The ring opening was followed by MALDI-TOF-MS, a mass increase of 40 Da (relative to the heterocyclised material) indicated successful incorporation of $^{18}$O into both carbonyl positions (Fig. 3.8). Once ring opening was complete, the sample was extensively dialysed into gel filtration buffer to remove all traces of acid and $H_2{}^{18}O$. Unexpectedly the sample formed a gel during dialysis. Full solubilisation of the material required the addition of 8 M urea, which was subsequently removed via passage over a gel filtration column, yielding pure $^{18}$O-PatE′.




(a)

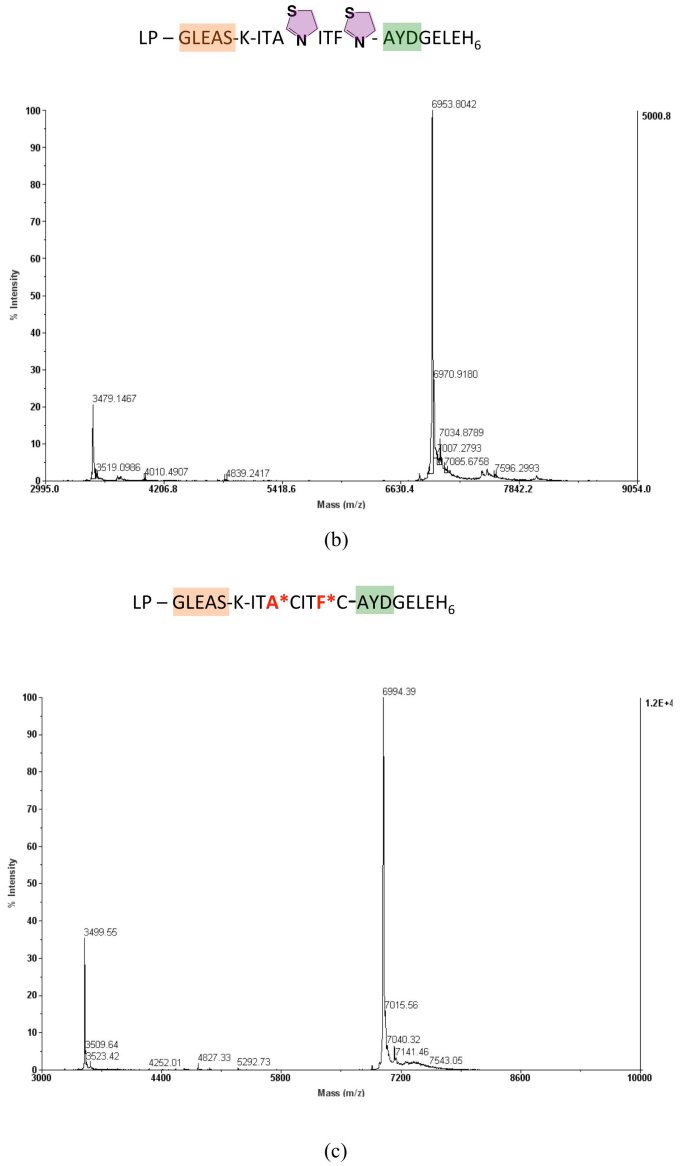



FIGURE 3.8: **Acidic ring opening of heterocyclised PatE′** (a) Mechanism of acidic ring opening of thiazoline heterocycles. (b) MALDI-TOF-MS of heterocyclised PatE′ (6953 Da). (c) MALDI-TOF-MS after ring opening by acidic $H_2{}^{18}O$ (6993 Da).

The $^{31}P$ NMR experiment was repeated using the selectively labelled$^{18}O$-PatE′ substrate. The result reflected that seen by Dunbar *et al.* in the BalhD study, whereby the heavy oxygen from the substrate carbonyl is incorporated into phosphate, as identified by an isotope shift of the $P_i$ peak in roughly 60:40 $^{16}O$ : $^{18}O$[50] ratio. Once again AMP and $PP_i$ were also detected in the $^{31}P$ NMR spectrum (Fig. 3.9). The same $^{16}O/^{18}O$ isotope shift was not observed for AMP, providing further evidence against an adenylation mechanism (Fig. 3.9), while the $PP_i$ peak appears to have broadened (Fig.

3.9). Addition of pyrophosphotase to the NMR tube increased the total amount of $P_i$ 2.7-fold (eliminating $PP_i$). Although the close proximity of the $^{16}O/^{18}O$ peaks results in some overlap, making it difficult to accurately integrate each peak, the ratio of $^{16}O/^{18}O$ $P_i$ did not appear appreciably altered (Fig. 3.9). If $PP_i$ production only arose due to background hydrolysis of ATP by LynD, then the $PP_i$ would only contain $^{16}O$. In this instance, addition of pyrophosphatase would only increase the $^{16}O$ $P_i$ peak, but not the $^{18}O$. The contrary increase of both the $^{16}O$ and $^{18}O$-$P_i$ peaks equally, is consistent with the presence of $^{18}O$ in $PP_i$.

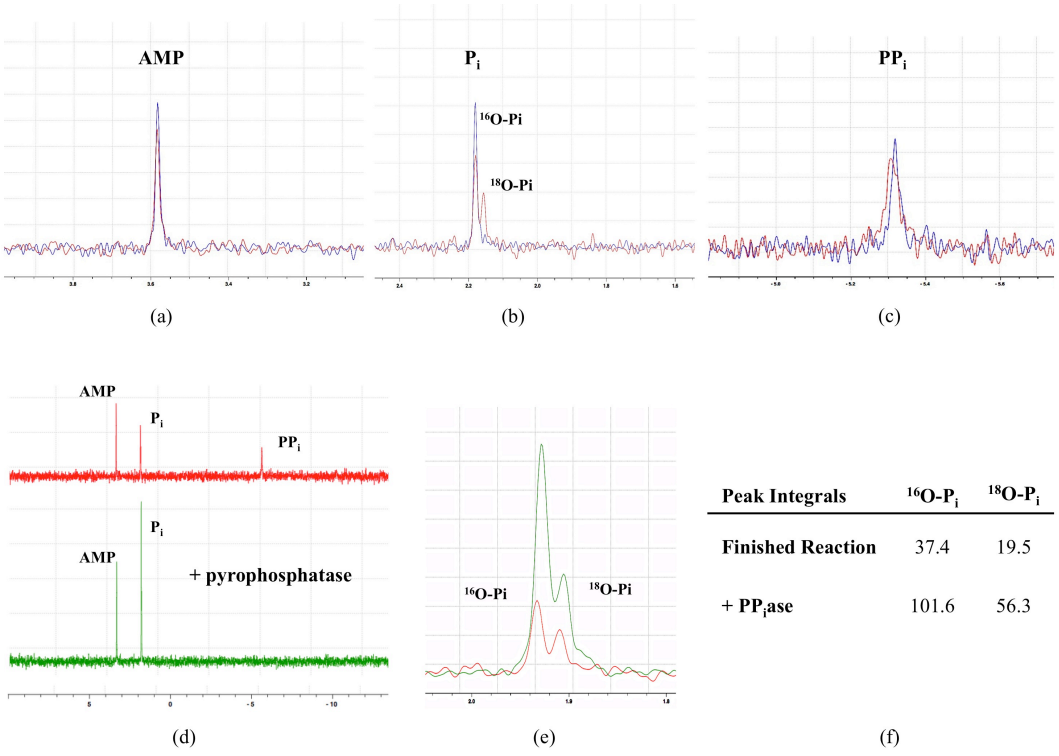



FIGURE 3.9: **$^{31}$P NMR spectrum of $^{18}$O-PatE′** $^{31}$P NMR spectrum of LynD + ATP + $^{16}$O-PatE′ (blue) overlaid with $^{31}$P NMR spectrum of LynD + ATP + $^{18}$O-PatE′ (red). (a) Close up of the AMP peak. (b) Close up of the $P_i$ peak; the peak is split in two because of the small but detectable chemical shift difference between $^{16}$O-$P_i$ and$^{18}$O-$P_i$.(c) Close up of the $PP_i$ peak; the $PP_i$ peak appears broader in the presence of $^{18}$O-PatE′. (d) $^{31}$P NMR spectrum of LynD + ATP + $^{18}$O-PatE′ before (red) and after (green) the addition of pyrophosphatase. After the addition of pyrophosphatase the $PP_i$ peak disappears and the $P_i$ peak grows in intensity. (e) Close up of the $P_i$ peak before (red) and after (green) the addition of pyrophosphotase. (f) Integration of $^{16}$O-$P_i$ and$^{18}$O-$P_i$ peaks, showing the proportion of each species is approximately the same following incubation with pyrophosphatase.

To provide further insight, we purified $PP_i$ from the $^{18}$O-PatE′ heterocyclisation reaction

mixture using anion exchange chromatography and analysed the purified sample using LC-ESI-MS (Fig. 3.10). These data revealed $^{16}O$ and $^{18}O$-$PP_i$ in approximately equal amounts ( m/z = 176.942 and 178.9466 respectively; Fig. 3.10). The lower than expected $^{18}O$ : $^{16}O$ ratio could be explained by $H_2{}^{16}O$ exchange with $PP_i$ under the LC-ESI-MS conditions, which has previously been seen for $P_i$ [102].

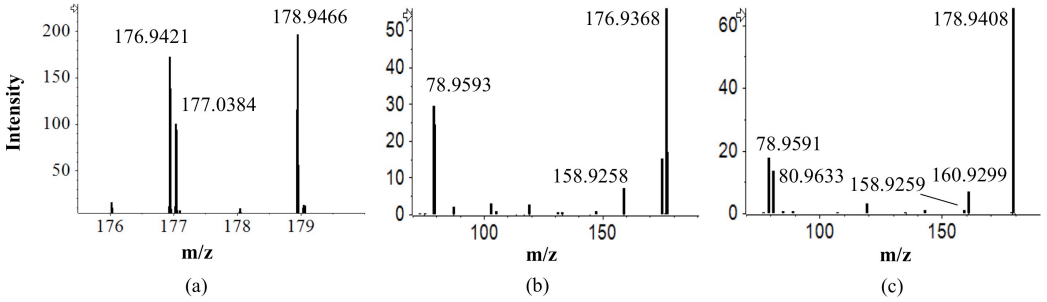



FIGURE 3.10: **MS of $PP_i$ purified from heterocyclisation reaction with $^{18}O$-PatE′** (a) $^{16}O$ and $^{18}O$-$PP_i$ (m/z = 176.9421 and 178.9466 respectively) are observed in roughly equal amounts. The peak at m/z = 177.0387 is a contaminant (formula $C_6H_{10}O_4P$ (the $^{12}C/^{13}C$ isotope pattern is seen)) and is not related to $PP_i$ and is observed in the $^{16}O$-$PP_i$ control. (b) MSMS fragmentation of $PP_i$ showing the typical fragmentation pattern: loss of water (m/z = 158.9258) and fragmentation into $PO_3^-$ (m/z = 78.9591). (c) MSMS fragmentation of $PP_i$ purified from the $^{18}O$-PatE′ heterocyclisation reaction. Loss of $H_2O$ and $H_2{}^{18}O$ (m/z = 160.9299 and 158.9259 respectively) and fragmentation into $PO_2{}^{18}O^-$ (m/z = 80.9633) indicates $^{18}O$ is incorporated in $PP_i$ during heterocyclisation

.

However the identification of $^{18}O$-$PP_i$ by LC-ESI-MS, confirmed by MSMS (Fig. 3.10) complements our $^{31}P$ NMR data implying $^{18}O$ from the carbonyl of the peptide substrate is transferred onto $PP_i$ during turnover. A feasible explanation of these data is a pyrophosphorylation mechanism, where the hemiorthoamide attacks the β-phosphate, breaking the α-β-phosphate bond to give AMP and $^{18}O$ pyrophosphorylated hemiorthoamide, which subsequently eliminates $^{18}O$-$PP_i$ (Fig. 3.11). The $^{18}O$-$PP_i$ then breaks down spontaneously under the reaction conditions to give $^{18}O$-$P_i$ and $^{16}O$-$P_i$.

Although less common than a kinase mechanism, precedent for such a mechanism in nature exists, with the best-known example of an enzyme of this class being ribose-phosphate-diphosphokinase, which catalyses the pyrophosphorylation of the 1′OH of ribose 5-phosphate [103]. Arguing against this hypothesis is the lack of structural homology between LynD and ribose-phosphate diphosphokinase. Furthermore, in all

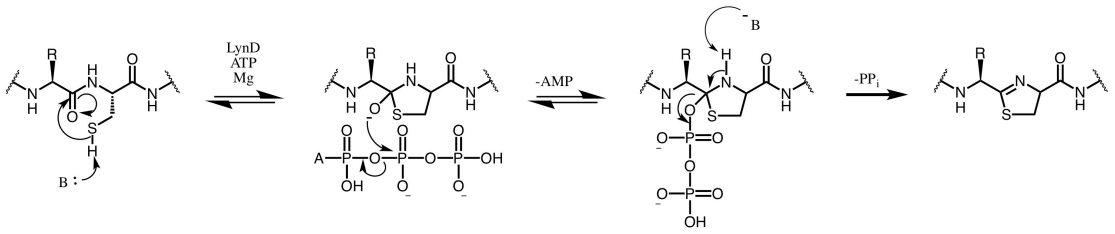
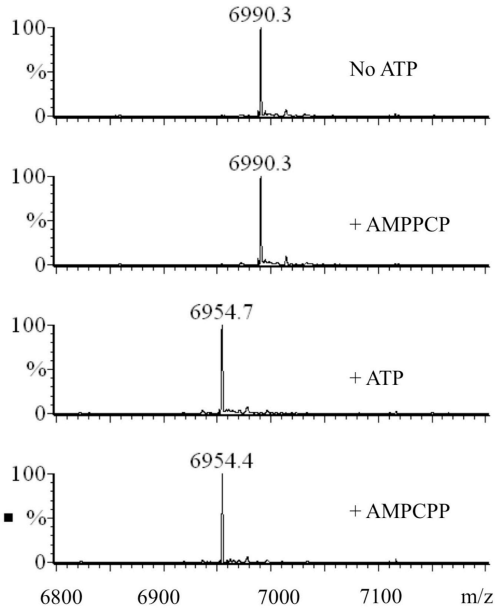


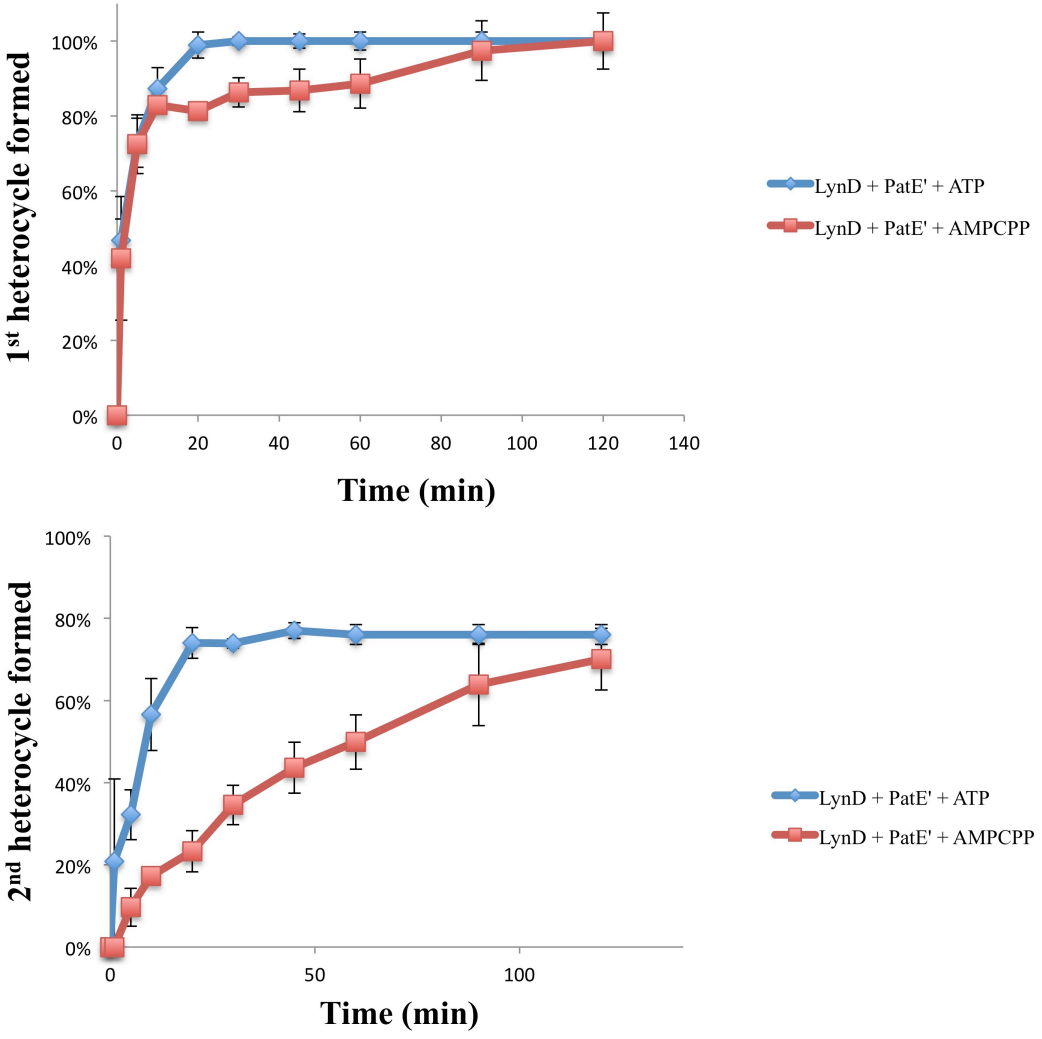

FIGURE 3.11: **Plausible pyrophosphorylation mechanism for heterocyclisation** In a reversible step a hemiorthoamide intermediate is formed. The hemiorthoamide intermediate subsequently attacks ATP at the β-phosphate, cleaving the α-β-phosphate bond and AMP is released. In an irreversible step the $PP_i$ is eliminated from the activated hemiorthoamide intermediate, yielding the heterocyclised product and $PP_i$.

our LynD structures the nucleotide is oriented such that the γ-phosphate is still more accessible to the hemiorthoamide intermediate than the β-phosphate (Fig. 3.4).

To test the hypothesis of a pyrophosphorylation mechanism, the ability of LynD to facilitate heterocyclisation in the presence of non-hydrolisable methylene bridged ATP analogues was investigated. Incubation of LynD with β,γ-methylene-ATP (AMPPCP) completely abolishes turnover, while α,β-methylene-ATP (AMPCPP) supports the first heterocyclisation event with a rate comparable to that of ATP, but is significantly slower than ATP for the second heterocyclisation event (Fig. 3.12). These data disprove the pyrophosphorylation mechanism as proposed (Fig. 3.11). The activity of the α,β-methylene bridged analogue implies that scission of the α,β-phosphate bond is not a requirement for turnover, while the inactivity of the β,γ analogue confirms that heterocyclisation must proceed via attack of hemiorthoamide at the γ position of ATP. A kinase mechanism is consistent with the observed activity of these ATP analogues, but does not account for the incorporation of $^{18}O$ in $PP_i$.


6990.3
100
%
0
No ATP
6990.3
100
%
0
+ AMPPCP
6954.7
100
%
0
+ ATP
6954.4
100
%
0
+ AMPCPP
6800
6900
7000
7100
m/z

(a)




(b)

FIGURE 3.12: **The effect of ATP analogues on substrate processing** (a) LC-ESI-MS of PatE′ (m/z = 6990) + LynD + ATP analogues. Complete substrate processing is evident by a loss of 2 water molecules ( - 36 Da, m/z = 6954). (b) Relative rates of heterocyclisation reaction between PatE′ and LynD + 5 mM ATP (blue) and PatE′ and LynD + 5 mM AMPCPP (red) analysed by MALDI-TOF-MS. The top graph shows the time taken to form the 1st heterocycle, and the bottom graph shows the time taken to form the 2nd heterocycle. Each experiment was set up in triplicate and each measurement was repeated three times; thus each time-point is an average of nine measurements. Errors are plotted as +/- 1 s.d..

Rationalising all the data leads us to propose what we believe to be a novel mechanism of substrate activation: we suggest that the reacting intermediate is first activated via phosphorylation, and the resulting *O*-phosphorylated hemiorthoamide subsequently attacks ADP, which is activated for cleavage by the $Mg^{2+}$ ions coordinated at the active

site, to yield AMP and O-$PP_i$ (Fig. 3.13). This second step, is not an occasional side reaction since ADP is not detected and $^{18}O$ is transferred into $PP_i$ efficiently. An immediate shortfall of this proposal is the heterocyclisation activity in the presence of AMPCPP, which should preclude the second phosphorylation. However this second phosphorylation event could be achieved by release of the AMPCPP hydrolysis product AMPCP (ADP analogue), followed by binding a new molecule of AMPCPP to supply a second $\gamma$-phosphate. Release and rebinding of the cofactor in this way could explain the slower rate of heterocyclisation observed for AMPCPP compared with ATP (Fig. 3.12). Unfortunately, a non-negligible rate of background hydrolysis of the nucleotide has thus far precluded our attempts to accurately perform the stoichiometry experiments necessary to confirm this hypothesis. Further work towards testing this mechanism, and attempts to unify all the observed data are ongoing.

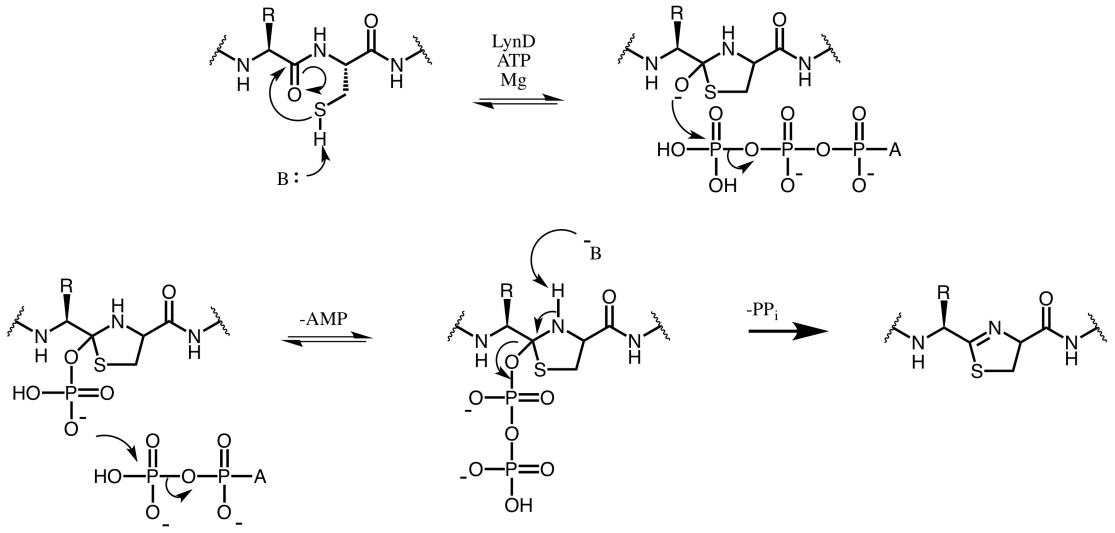



FIGURE 3.13: **Proposed heterocyclisation mechanism** The hemiorthoamide attacks the $\gamma$-phosphate of ATP, cleaving the $\beta$,$\gamma$-phosphate bond giving ADP. The ADP is not released from the enzyme, and is attacked by the phosphorylated hemiorthoamide intermediate at the $\beta$-phosphate to give AMP. In an irreversible step the $PP_i$ is eliminated from the activated hemiorthoamide intermediate, yielding the heterocyclised product and $PP_i$.

### 3.3.5 Leader Peptide Recognition

In all three LynD complex structures, only residues Q21-G35 of the leader peptide of PatE′ are ordered, and are located at the interface between domain 1 from one LynD monomer and domain 3 from another (Fig. 3.14). This ordered region of the leader

peptide in the structures corresponds with the residues thought to govern substrate recognition, as identified by multiple biochemical studies, providing rational to the 'minimal leader'[41,51,52]. Residues Q21-S23 adopt a coil, S24-E28 a helical turn, S30-A33 a β-strand, and L34-G35 a coil. The β-strand adds to the antiparallel β-sheet in domain 1 (Fig. 3.14). This extended β-sheet interaction is similar to that observed in the lantibiotic dehydratase substrate complex[104]. In addition to the β-sheet hydrogen bonds, the chain of PatE′ E28 and E31 make hydrogen bonds and salt bridges with the main chain of LynD Q37 and the side chains of Y39 and R74. Residues L26 and L29 of PatE′ are found in hydrophobic pockets of domain 1 making van der Waals contacts with Y67 and Y70. The leader peptide makes contacts with domain 3 of LynD through a salt bridge between PatE′ E32 and LynD R399 and hydrophobic interactions between PatE′ L34 with V217 and L398 of LynD. The specific contacts made between the leader peptide and domain 3 of LynD are responsible for the ordering of a region that was not observed in the apo structure of TruD[51].

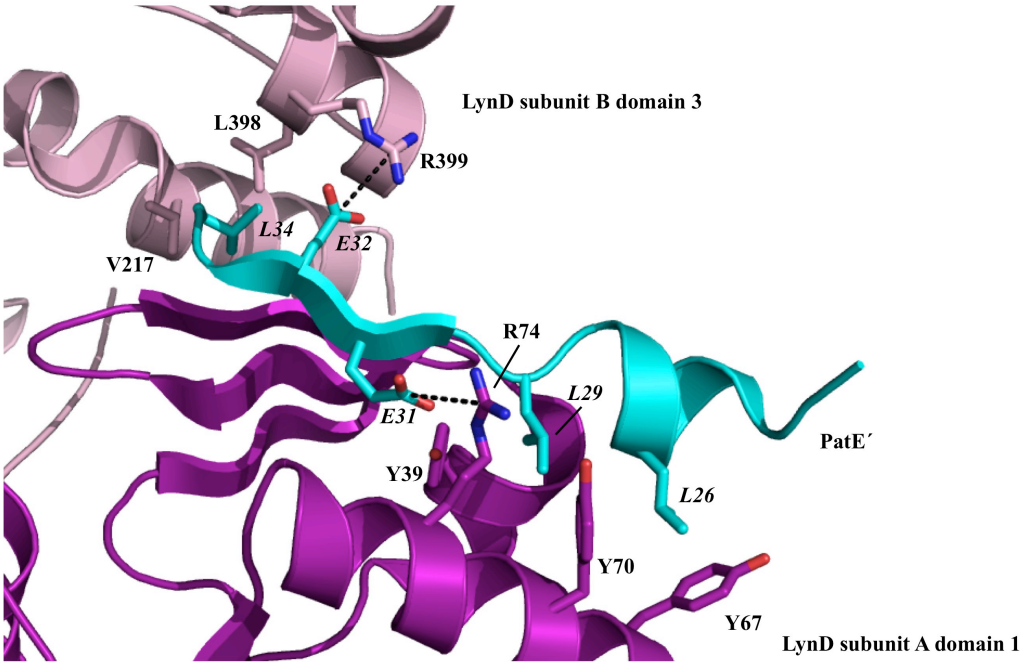



FIGURE 3.14: **Leader peptide binding by LynD** Ordered residues of PatE′ are shown as a cyan cartoon and subunit A and subunit B of LynD are shown as purple and pink cartoons respectively. Residues directly involved in binding are shown as sticks. Numbering of PatE′ residues is shown in italics for clarity. Salt bridge interactions are shown with dashed lines.

To validate the importance of the residues identified in the crystal structures in substrate

binding, a series of mutants of LynD and PatE′ were made and their binding interactions measured using ITC (Fig. 3.15). Furthermore the effect of each mutant on substrate processing was monitored using MALDI-TOF-MS. A summary of the binding and activity data is shown in Table 3.2.

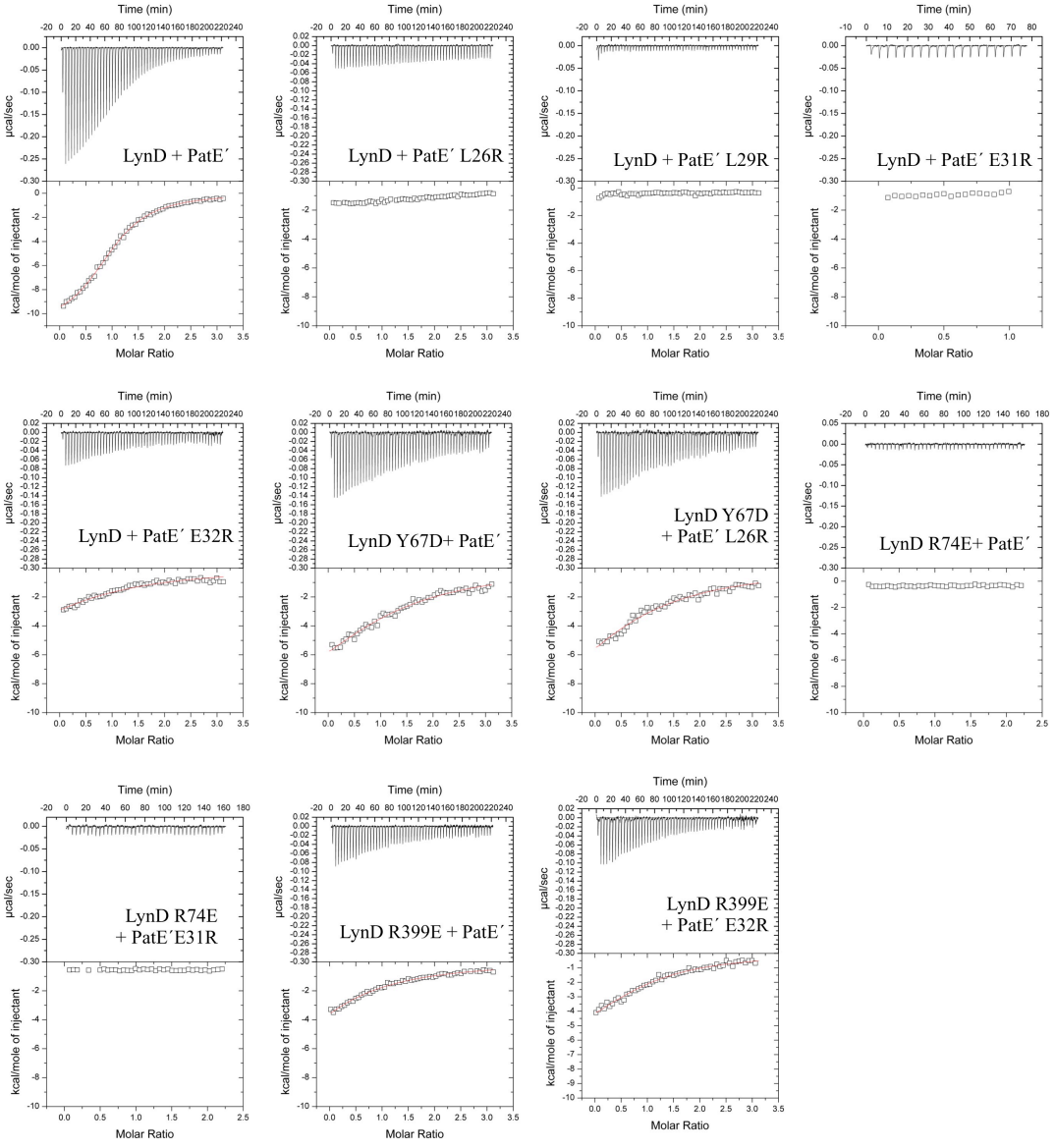



FIGURE 3.15: **LynD substrate binding** ITC data for PatE′ and PatE′ mutants titrated into LynD and LynD mutant solutions. The top panel shows raw data representing heat evolved in response to injections, the bottom panel shows the integrated heats of injections (□) and the best fit (-) to the one-site model (*Origin*).

LynD binds PatE′ with a $K_D$ = 1.49 µM (Fig. 3.15). This is consistent with the low µM binding observed between TruD and PatE2[51]. The PatE′ L26R, L29R and E31R

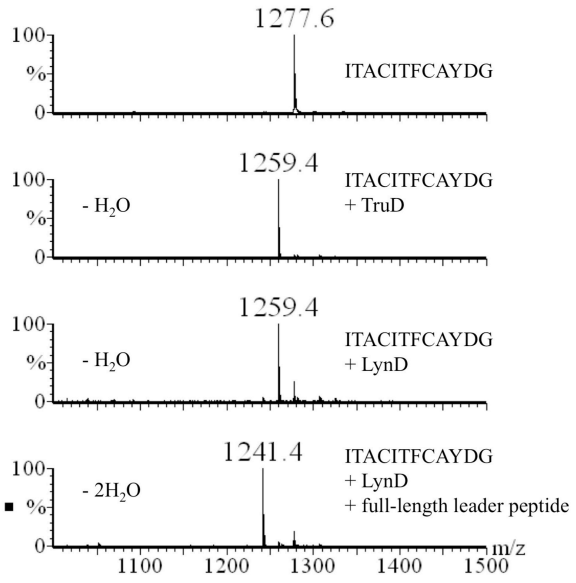
TABLE 3.2: LynD and PatE′ mutations and their effects on binding and processing.

| **Enzyme** | **Peptide** | **$K_D$ (μM)** | **No. of heterocycles (PatE′)** |
|---|---|---|---|
| LynD | PatE′ | 1.49 | 2 |
| LynD | PatE′$^{L26R}$ | No binding | 2 |
| LynD | PatE′$^{L29R}$ | No Binding | 1 and 2 |
| LynD | PatE′$^{E31R}$ | No Binding | 1 and 2 |
| LynD | PatE′$^{E32R}$ | 18.35 | 2 |
| LynD$^{Y67D}$ | PatE′ | 12.92 | 2 |
| LynD$^{Y67D}$ | PatE′$^{L26R}$ | 15.22 | 2 |
| LynD$^{R74E}$ | PatE′ | No binding | 2 |
| LynD$^{R74E}$ | PatE′$^{E31R}$ | No binding | 1 and 2 |
| LynD$^{R399E}$ | PatE′ | 10.74 | 2 |
| LynD$^{R399E}$ | PatE′$^{E32R}$ | 8.55 | 2 |

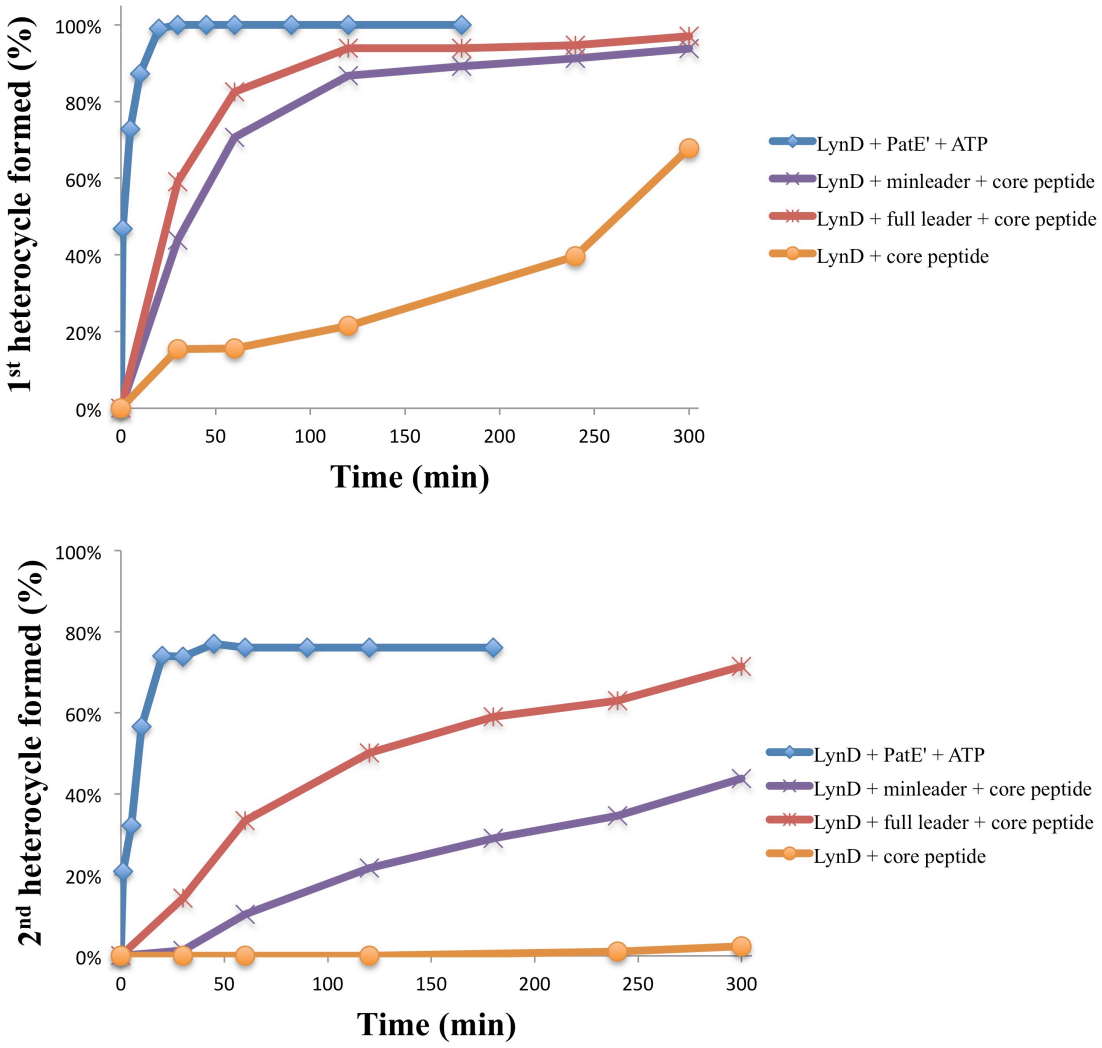
mutations reduce binding to wild-type LynD below the detection limits of ITC, with the later two mutations affecting processing by LynD under our standard reaction conditions, forming a mixture of 1 and 2 heterocycles (Fig. 3.15, Table 3.2). PatE′$^{E32R}$ shows weaker binding to LynD, with a $K_D$ = 18.35 μM, but is completely processed by LynD under our standard conditions (Fig. 3.15, Table 3.2). The LynD Y67D, R74E and R399E mutations all reduce binding affinity to PatE′, but retain catalytic activity (Fig. 3.15, Table 3.2). In pairing the LynD mutants with their corresponding PatE′ mutants, such that the interacting enzyme and substrate residues again have opposite charge (or charge introduced in the case of LynD$^{Y67D}$ and PatE′$^{L26R}$) we sought to investigate whether binding could be restored. However these coupled mutations had little or no effect on the binding affinity between enzyme and substrate, compared with the single mutation experiments. Interestingly LynD$^{R74E}$ struggled more to accept PatE′$^{E31R}$ as a substrate than wild-type PatE′, producing a mixture of 1 and 2 heterocycles within the time-frame of our assay (Fig. 3.15, Table 3.2). These data confirm the importance of the residues identified in the crystal structures in substrate recognition.

### 3.3.6 Activation of LynD by the Leader Peptide

The cyanobactin heterocyclases TruD and PatD have been shown to process leaderless peptide substrates *in vitro*, but inefficiently; they cannot process all of the potentially heterocyclisable residues within the peptide[41,51]. Thus the leader peptide was deemed important for full activity of the heterocyclase, and not a prerequisite for heterocyclisation. Using a short, leaderless peptide substrate ITACITFCAYDG (mimicking the core peptide and C-terminal macrocyclisation element of PatE′), LynD catalysed a single cyclodehydraton reaction after overnight incubation, which is consistent with previous reports[51] (Fig. 3.16). An elegant study by Y. Goto *et al.* had shown that full processing of all heterocyclisable residues in a short peptide substrate by PatD can be recovered through addition of exogenous leader peptide[41]. Here, we show that addition of full-length PatE′4 leader (up to but not including the core peptide) and minimal leader (LAELSEEAL) peptides both restore processing by LynD, converting both cysteine residues in ITACITFCAYDG to thiazolines, but at a slower rate compared to LynD and full-length PatE′ (Fig. 3.16). The increase in activity is more pronounced for the full-length leader peptide than for the minimal leader peptide (Fig. 3.16).


1277.6
100
%
0
ITACITFCAYDG
1259.4
100
%
0
- H2O
ITACITFCAYDG
+ TruD
1259.4
100
%
0
- H2O
ITACITFCAYDG
+ LynD
1241.4
100
%
0
- 2H2O
ITACITFCAYDG
+ LynD
+ full-length leader peptide
m/z
1100
1200
1300
1400
1500

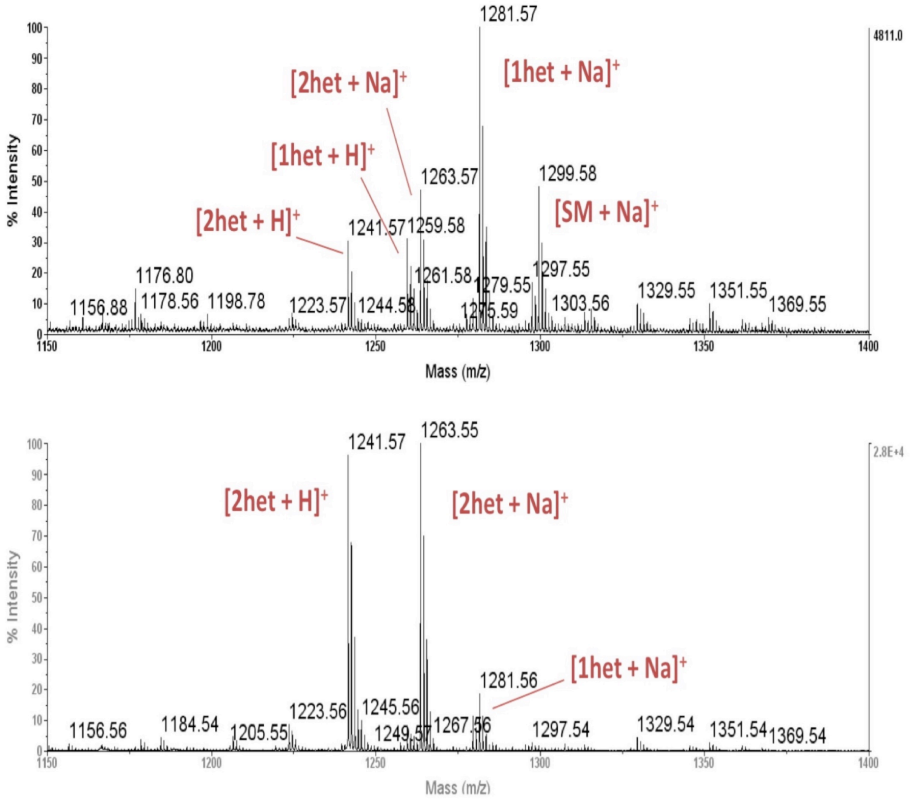
(a)



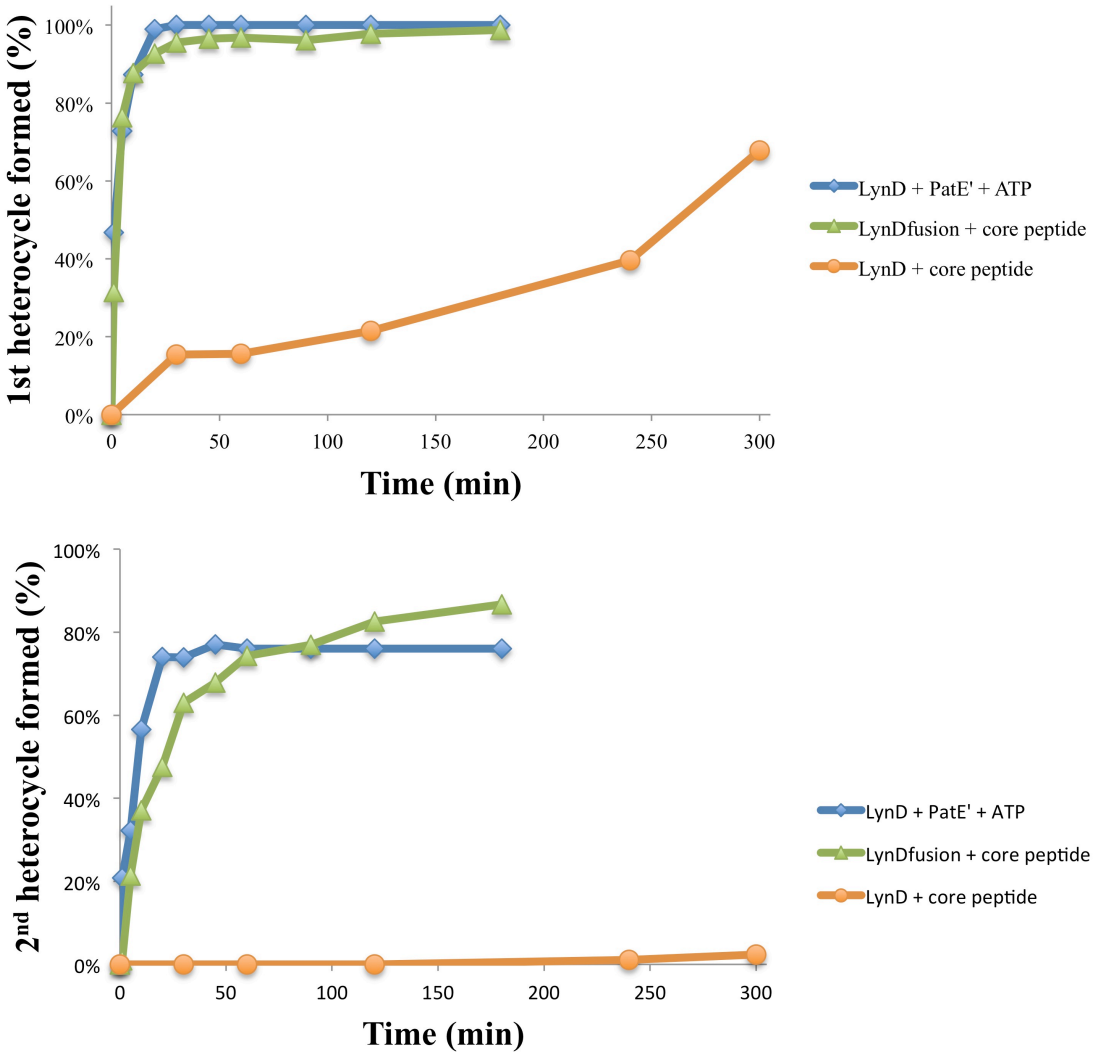

(b)

FIGURE 3.16: **The roles of substrate leader in promoting catalysis** (a) LC-ESI-MS analysis of ITACITFCAYDG (m/z = 1277.6) incubated with TruD, LynD +/- full-length leader peptide overnight. When the substrate peptide is incubated with just TruD or LynD only 1 heterocycle is formed (m/z = 1259.4). Addition of exogenous full-length leader peptide restores processing by LynD, such that 2 heterocylces are formed following an overnight incubation. (b) Relative rates of various leader peptide-activated heterocyclisation reactions analysed by MALDI-TOF-MS: LynD + PatE′ (blue), LynD + minimal leader + ITACITFCAYDG (purple), LynD + full-length leader + ITACITFCAYDG (red) + LynD + ITACITFCAYDG (orange). The top graph shows the time taken to form the 1st heterocycle, and the bottom graph shows the time taken to form the 2nd heterocycle. Each experiment was set up in triplicate and each measurement was repeated three times; thus each time-point is an average of nine measurements. Errors are plotted as +/- 1 s.d..

Guided by structural analysis, residues Q21 to A37 of the leader peptide were fused to the N-terminus of LynD to construct a self-activating enzyme (AcLynD). A six-pair 'GA' linker was introduced between the last residue of the leader peptide and the first residue of LynD to promote correct folding of the leader peptide into the enzyme binding site. This engineered enzyme expressed at a higher yield than native enzyme, and processed both cysteine residues in ITACITFCAYDG substrate at a similar rate to wild-type LynD with full-length PatE′ (Fig. 3.17).


1281.57
4811.0
[2het + Na]+
[1het + Na]+
[1het + H]+
1263.57
1299.58
[SM + Na]+
[2het + H]+
1241.57
1259.58
1261.58
1297.55
1176.80
1279.55
1329.55
1351.55
1156.88
1178.56
1198.78
1223.57
1244.58
1303.56
1369.55
% Intensity
Mass (m/z)
1241.57
1263.55
2.8E+4
[2het + H]+
[2het + Na]+
[1het + Na]+
1281.56
1223.56
1245.56
1156.56
1184.54
1205.55
1249.57
1267.56
1297.54
1329.54
1351.54
1369.54
% Intensity
Mass (m/z)

(a)

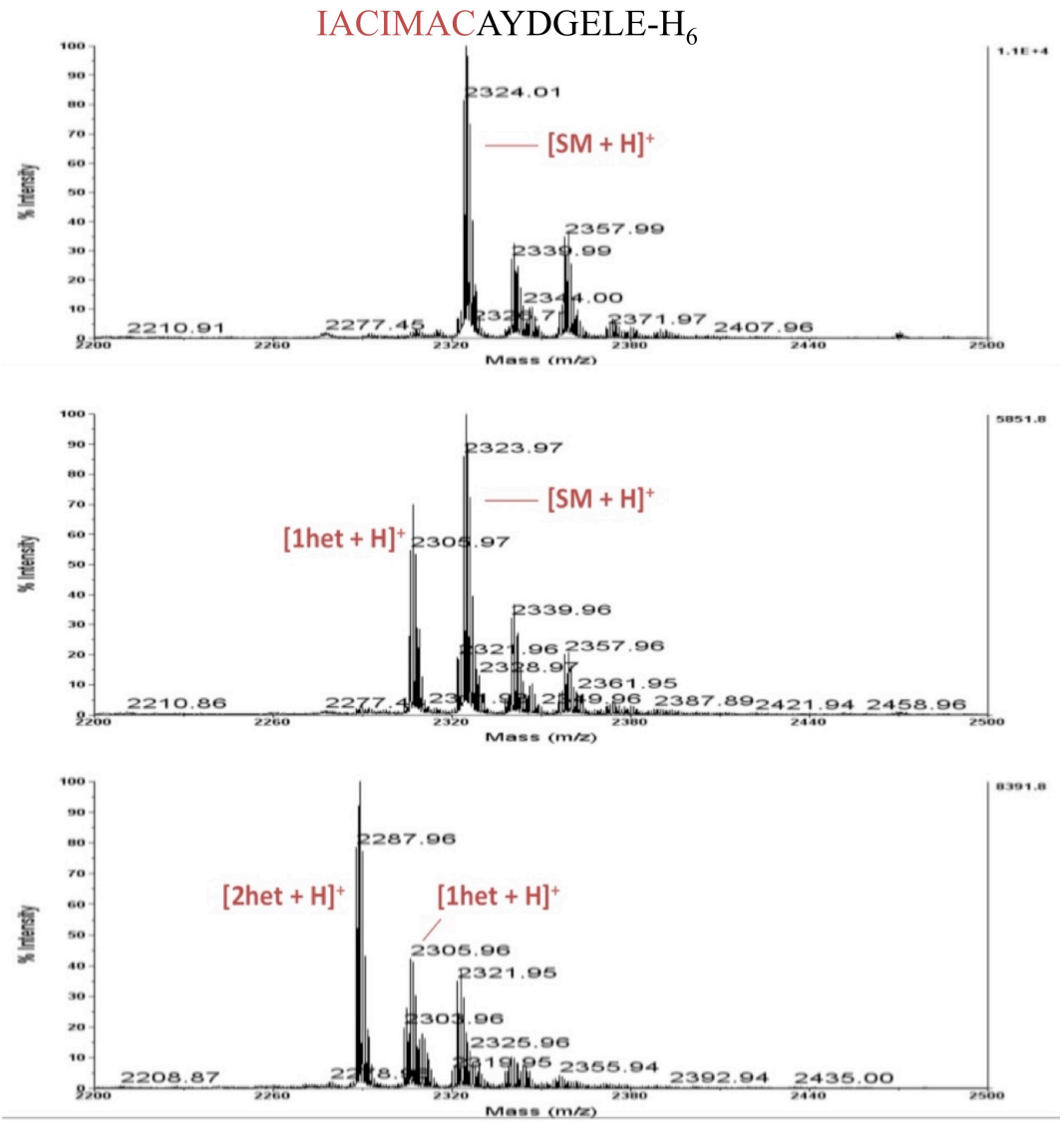



(b)

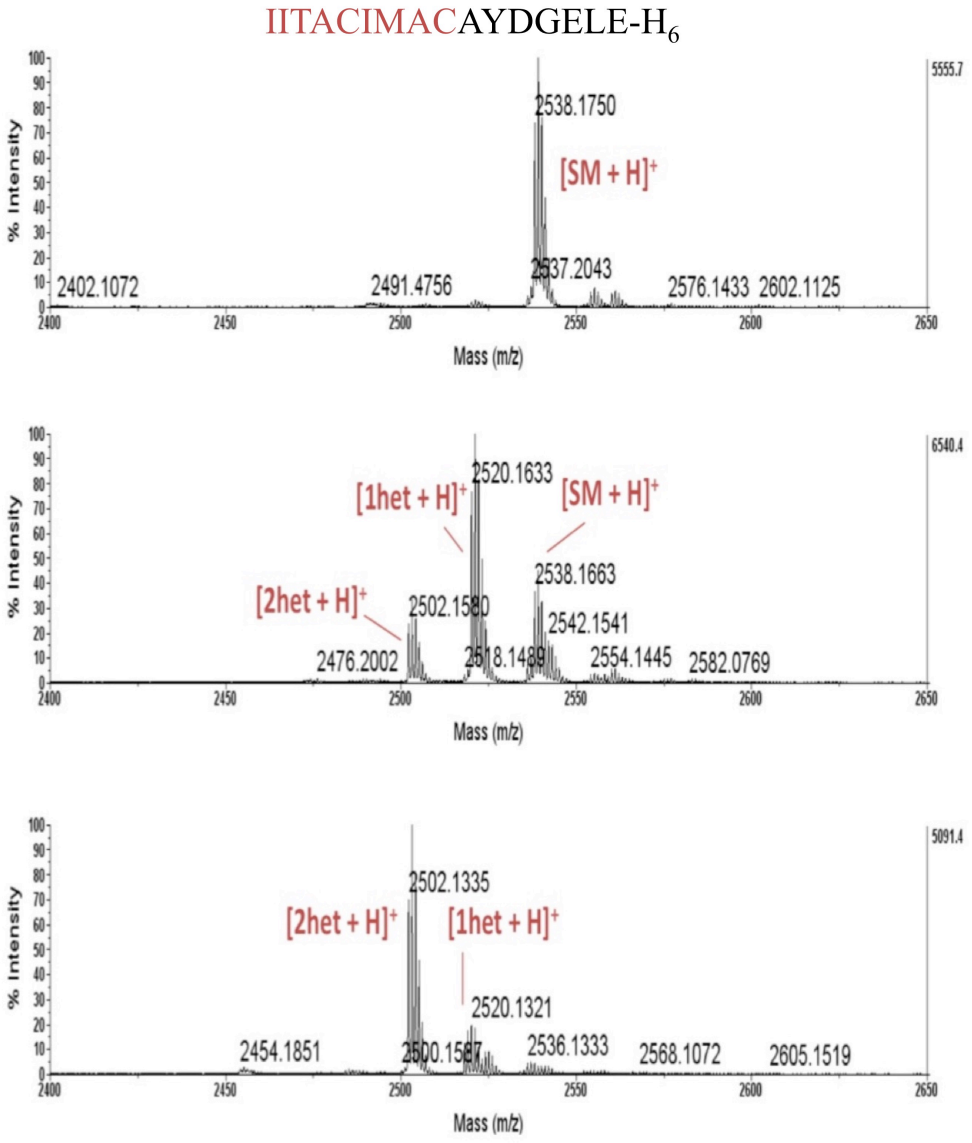
FIGURE 3.17: **Activity of AcLynD on ITACITFCAYDG** (a) MALDI-TOF-MS analysis of ITACITFCAYDG (m/z = 1277.6) incubated with AcLynD overnight. The substrate is fully processed by AcLynD to give a product containing 2 heterocycles. (b) Relative rate of AcLynD + ITACITFCAYDG (green) compared with LynD + PatE′ (blue) and LynD + ITACITFCAYDG (orange) analysed by MALDI-TOF-MS. The top graph shows the time taken to form the 1st heterocycle, and the bottom graph shows the time taken to form the 2nd heterocycle. Each experiment was set up in triplicate and each measurement was repeated three times; thus each time-point is an average of nine measurements. Errors are plotted as +/- 1 s.d..

To investigate whether the AcLynD exhibits a high substrate tolerance as seen for the wild-type enzyme, a variety of short test peptide substrates were generated in house, by mutating the core peptide of PatE′ and removing the leader peptide with trypsin prior to the heterocyclisation reaction. In this way three test peptides were produced,

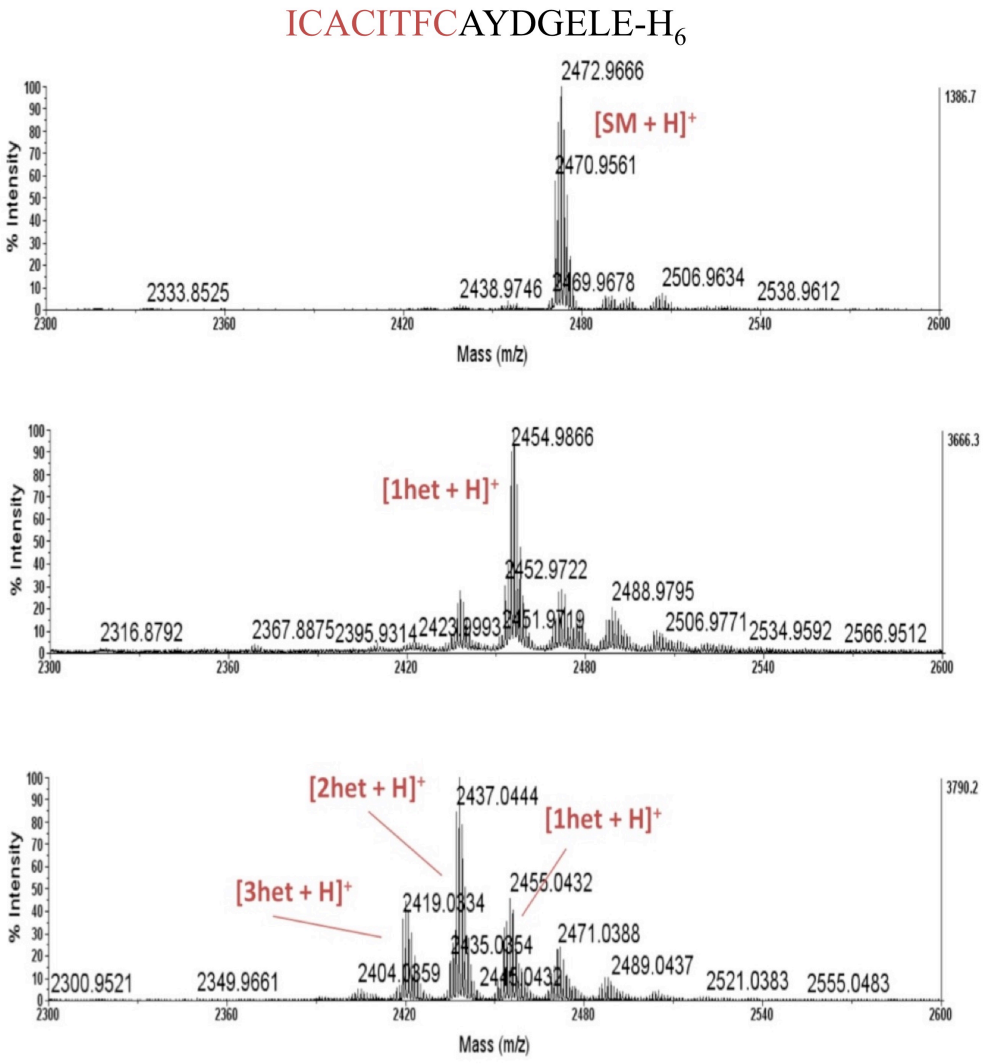
which contain a core peptide sequence, the macrocyclisation signature (AYDG) and a C-terminal extension (ELEHHHHHH (ELE-$H_6$)) derived from the original PatE′ sequence: a peptide with a seven residue core peptide (IACIMACAYDGELE-$H_6$), a peptide with a nine residue core peptide (IITACIMACAYDGELE-$H_6$) and a peptide with three cysteine residues within the core peptide (ICACITFCAYDGELE-$H_6$). The three test peptides were incubated with AcLynD and native LynD under standard conditions. Both IACIMACAYDGELE-$H_6$ and IITACIMACAYDGELE-$H_6$ were completely processed by AcLynD, while ICACITFCAYDGELE-$H_6$ gave a mixture of 3 and 2 heterocycles (Fig. 3.18). AcLynD showed a significant increase in activity compared with native LynD, which only generated 1 heterocycle in all three substrates (Fig. 3.18). Although ICACITFCAYDGELE-$H_6$ was not processed to completion under standard conditions, indicating the need for optimisation, these data emphasise the advantageous ability of AcLynD to process short leaderless substrates with a much greater efficiency than the native enzyme.


IACIMACAYDGELE-H6
1.1E+4
2324.01
[SM + H]+
2357.99
2339.99
2344.00
2210.91
2277.45
2371.97
2407.96
% Intensity
Mass (m/z)
5851.8
2323.97
[SM + H]+
[1het + H]+
2339.96
2357.96
2361.95
2210.86
2387.89
2421.94
2458.96
8391.8
2287.96
[2het + H]+
[1het + H]+
2305.96
2321.95
2325.96
2208.87
2355.94
2392.94
2435.00

(a)


IITACIMACAYDGELE-H6
% Intensity
Mass (m/z)
2538.1750
[SM + H]+
2537.2043
2402.1072
2491.4756
2576.1433
2602.1125
5555.7
2520.1633
[1het + H]+
[SM + H]+
2538.1663
[2het + H]+
2502.1580
2542.1541
2476.2002
2518.1489
2554.1445
2582.0769
6540.4
2502.1335
[2het + H]+
[1het + H]+
2520.1321
2454.1851
2500.1587
2536.1333
2568.1072
2605.1519
5091.4

(b)




(c)

FIGURE 3.18: **Activity of AcLynD with various short leaderless substrate peptides** MALDI-TOF-MS analysis of AcLynD + (a) IACIMACAYDGELE-$H_6$ (b) IITACIMACAYDGELE-$H_6$ and (c) ICACITFCAYDGELE-$H_6$. The top panel shows the substrate peptide on its own, the middle panel shows the substrate peptide processed by LynD and the bottom panel shows the substrate peptide processed by AcLynD.

AcLynD can be thought of as an 'activated' enzyme, and its activity compared with wild-type enzyme is summarised by the cartoon schematic seen in Fig. 3.19.

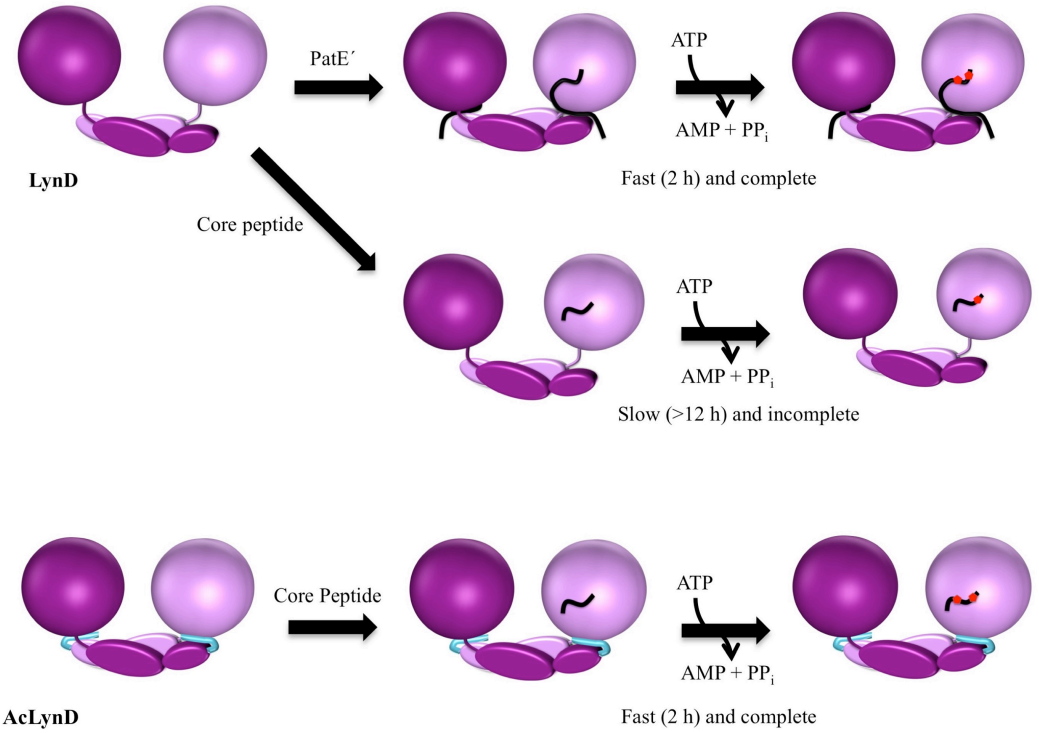



FIGURE 3.19: **Cartoon schematic of LynD and AcLynD** The LynD dimer undergoes a conformational rearrangement (purple and pink) following binding of PatE′ (black) generating an 'active' enzyme. In the absence of a leader peptide the conformational rearrangement is not possible and heterocyclisaton of the core peptide is inefficient. Covalent attachment of the leader peptide to the enzyme (cyan) in AcLynD mimics the 'activated' native LynD, allowing efficient wild-type-like processing of the core peptide.

## 3.4 Conclusions

The cyanobactin heterocyclase enzymes (D enzymes) are responsible for the ATP-dependent post-translational conversion of cysteine (and in some cases serine and threonine) residues to thiazoline (and oxazoline) heterocycles within the cyanobactin precursor peptide[40]. A key feature of these enzymes is their ability to separate substrate recognition from catalysis, allowing them to process numerous precursor peptides containing varied core peptide sequences, ultimately leading to diverse products, as long as recognition elements within the leader peptide are conserved[51,52]. This low substrate specificity makes the D enzymes attractive enzymes for biotechnological development. The crystal structures of heterocyclase LynD, from the aestuaramide pathway (*Lyngbya sp.*), in complex with various substrate peptides and nucleotides were solved, allowing both the active site and the substrate recogniton site to be located.

These complex structures revealed the α-phosphate of ATP, bound in domain 3 (the conserved YcaO domain[49,101]) of LynD, is shielded from incoming substrate peptide, and thus heterocyclisation cannot proceed via an adenylation mechanism. While the exposed γ-phosphate supports a kinase mechanism, a kinase mechanism cannot explain the observation that LynD, as with TruD, hydrolyses ATP to AMP and $PP_i$ during turnover[51]. In an extension of the selective $^{18}O$ labelling studies of the LAP heterocyclas BalhD, $^{18}O$ from the carbonyl of the residue directly preceding the cyclisable-cysteine was shown to be efficiently transferred into $PP_i$, confirming $PP_i$ is a mechanistically relevant bi-product of heterocyclisation, and does not solely arise due to background hydrolysis of ATP. The non-hydrolysable ATP analogue AMPCPP supported catalysis but at a slower rate than that seen for ATP, however the AMPPCP analogue abolished catalysis, indicating heterocyclisation occurs through attack of the substrate of the γ-phosphate of ATP. A plausible mechanism, unifying all these data was proposed, where the cysteine thiol cyclises onto the neighbouring carbonyl carbon, creating a hemiorthoamide intermediate, which subsequently attacks the γ-phosphate of ATP at the active site. At this point the resulting ADP is not released from the enzyme active site, but attacked at the β-phosphate (or a second equivalent of nucleotide in the case of the AMPCPPP analogue) by the activated phosphorylated hemiorthoamide, creating a pyrophosphorylated hemiorthoamide, which eliminates to give the thiazoline product, and $PP_i$.

The substrate binding site was located at the interface of domain 1 from one monomer of LynD and domain 3 from the other, so the dimer is a functional requirement. Binding of the substrate brings about a significant conformational rearrangement of the enzyme and the holoenzyme dimer is packed together much more tightly compared with the apo TruD structure. Furthermore the rearrangement upon binding results in the ordering of a number of previously disordered residues within domain 3, including a helix-strand-helix motif that forms part of the entrance to the nucleotide binding site. As expected, substrate binding is dominated through interactions between LynD and the leader peptide, with only residues Q21-G35 of the leader peptide ordered in all the complex structures. The ordered region overlaps well with those residues thought to play an important role in substrate recognition as identified through numerous biochemical studies[41,51,52] and rationalises the previously described 'minimal leader'[51].

Similarly to TruD, LynD is capable of installing a single heterocycle, at the C-terminus

of a core peptide of a leaderless substrate. Upon addition of exogenous 'minimal' leader or full-length leader peptide, the processing ability of LynD is restored, cyclising both cysteine residues in the leaderless substrate. The efficiency of this *trans* activation of LynD towards leaderless substrates was significantly increased by fusing the important leader peptide residues to the N-terminus of LynD, creating an 'activated' enzyme: AcLynD. AcLynD has been shown to efficiently process multiple leaderless substrates, and at a rate comparable to that of the native enzyme with the full-length PatE′ substrate. The ability to dispose of the leader peptide is highly advantageous from a biotechnological perspective, as it allows for the synthesis of diverse cyanobactin analogues *in vitro*, starting from shorter, thus more economic substrates. Furthermore, shorter peptides are more amenable to chemical synthesis, allowing for the incorporation of non-natural amino acids, offering an explosion in diversity. Moreover AcLynD, and homologues thereof are potentially capable of installing heterocycles in substrates, unrelated to cyanobactins, vastly increasing their potential biotechnogical application.

## 3.5 Future Work

The heterocyclisation mechanism proposed in this chapter is the only feasible mechanism we can see that consolidates all the data acquired to date. However this mechanistic hypothesis remains to be tested. If the stoichiometry of both ATP and AMPCPP required for the heterocyclisation reaction can be accurately determined, this will address whether two phosphorylation events are necessary for heterocycle formation, and provide evidence that the proposed mechanism is plausible.

Expansion of the AcLynD technology to homologues able to efficiently modify serine and threonine residues has since been carried out, with AcMicD and AcPatD constructs having been cloned by Ms Ying Ge. Work towards optimising the activity, and investigation into the substrate tolerance of the AcD enzymes, and thus the potential limitations of the technology are ongoing.

# Chapter 4

# Investigation into Cyanobactin Epimerisation

## 4.1 Introduction

Biology is chiral. Amino acids (except glycine) and thus proteins are inherently chiral. It follows therefore, that the chirality of a drug molecule is critically important to its biological activity[105,106]. Despite the same chemical structure, stereoisomers of chiral drugs can exhibit significant differences in their pharmacological properties, as well as differences in the way they are metabolised[106]. Single enantiomer drugs can display enhanced target selectively, thus reducing off-target toxicity, compared with racemic mixtures.[105]. Most famously, the lack of enantiomeric purity of the sedative Thalidomide led to the unexpected and tragic teratogenic side effects, that has been described as a medical disaster[107]. Consequently, in the preparation of pharmaceutical compounds it is imperative that the stereochemistry of chiral centres of the final compounds are characterised, and when possible produced as single enantiomers, either during synthesis or as a result of purification[105,106].

Several cyanobactins have been confirmed to contain D-stereocentres, and although not proven, they are likely widely distributed throughout the family[45,61–66]. The conservation of these epimerised residues suggests they may be important for activity. Epimerisation during cyanobactin biosynthesis has been proposed to be a spontaneous process[45,67], although the involvement of enzymes in this process has not been formally

excluded. Furthermore, it is not known at what point during cyanobactin biosynthesis the epimerisation reaction occurs. If it is a spontaneous process, it would seem likely to occur on the macrocycle as a means to relieve steric pressure and form a thermodynamically stable conformation[67]. Addtionally D-stereocentres have not been observed in the related linear azoline containing peptides (LAPs)[20]. Cyanobactin products made *in vitro* give rise to CD-spectra that is identical to that of the natural products, and on this basis it was concluded epimerisation had occurred spontaneously under the reaction conditions used in the laboratory[47]. Nevertheless it is important to characterise the epimerisation reaction, to definitively address the uncertainties surrounding the reaction, and determine whether an epimerisation step needs to be built into future developments of the *in vitro* synthesis. Once the epimerisation reaction has been characterised, it may be possible to control the epimerisation of select residues, increasing the diversity accessible through the *in vitro* synthesis.

Epimerisation reactions at a carbon centre require hydrogen exchange and so can usually be studied using $^{1}$H NMR and related techniques. For epimerisation events where proton abstraction and addition is mediated via water, dissolving the molecule in $D_2O$ will result in a loss of proton peaks in the NMR spectrum as the labile hydrogen is replaced with solvent deuterium (spin = 1). Such an experiment is straightforward for small molecules, which do not contain many hydrogen atoms, and thus individual atoms can be monitored. However epimerisation of cyanobactins could arise directly after heterocyclisation of the 60 + residue precursor peptide. The derivatives of PatE used in the *in vitro* synthesis (PatE′) contain >450 hydrogen atoms. Scanning a $^{1}$H NMR spectrum for the loss of two, individual proton peaks resulting from an epimerisation reaction would be at best challenging, as can be seen in the $^{1}$H NMR spectrum of PatE′ (Fig. 4.1). Regions of characteristic chemical shifts have been highlighted.

In theory structural information of larger peptides, such as PatE′, can be resolved using 3D triple resonance ($^{1}$H, $^{13}$C, and $^{15}$N) NMR experiments[108]. By compiling this information, it should be possible to identify the αC protons as cross peaks in a $^{1}$H-$^{13}$C-HSQC spectrum, and subsequently observe whether they disappear due to exchange with solvent deuterium. The unfolded nature of PatE results in a narrow dispersion of cross peaks in 2D spectra, resulting in overlapping cross peaks[51], and so complicates the detection of the two peaks of interest. The AcD enzymes (chapter 3) are capable of processing short, leaderless substrates, which are easily accessible via chemical synthesis.

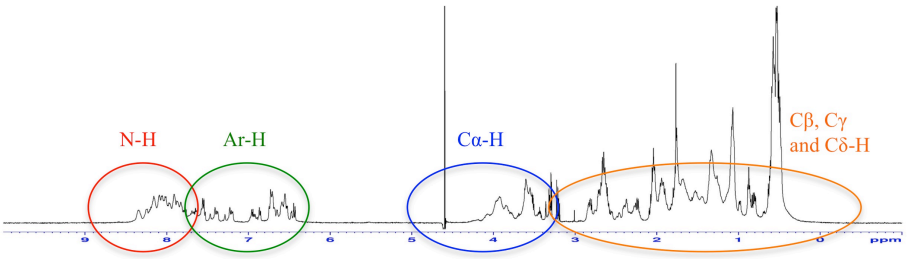



FIGURE 4.1: **PatE′ $^{1}$H NMR spectrum** PatE′ contains 478 hydrogens. The resolution is not sufficient to identify individual proton resonances.

Isotopically labelled amino acids can be incorporated into the peptide synthesis to generate starting material where the epimerisable residues are selectively labelled. Using $^{13}$C-alanine (A*) the selectively-labelled starting peptide ITA*CITA*CAYDGE was synthesised (BioSyn). This peptide will have only two peaks in a $^{1}$H-$^{13}$C-HSQC spectrum ($\alpha$CH and $\beta CH_3$) and so analysis of the epimerisation reaction is greatly simplified. A reaction scheme using the starting peptide ITA*CITA*CAYDGE can be seen in Fig. 4.2.

$^{1}$H-$^{13}$C-HSQC NMR of the substrate ITA*CITA*CAYDGE can be used to monitor epimerisation on the linear peptide following heterocyclisation by the AcD enzymes, and following macrocyclisation and oxidation, by PatGmac and ArtGox respectively, by conducting the enzymatic transformations in $D_2O$ buffer. Loss of the $\alpha$CH cross peak in the $^{1}$H-$^{13}$C-HSQC spectrum confirms H/D exchange has occurred, which is indicative of epimerisation. The presence of the $\beta CH_3$ provides an internal control.

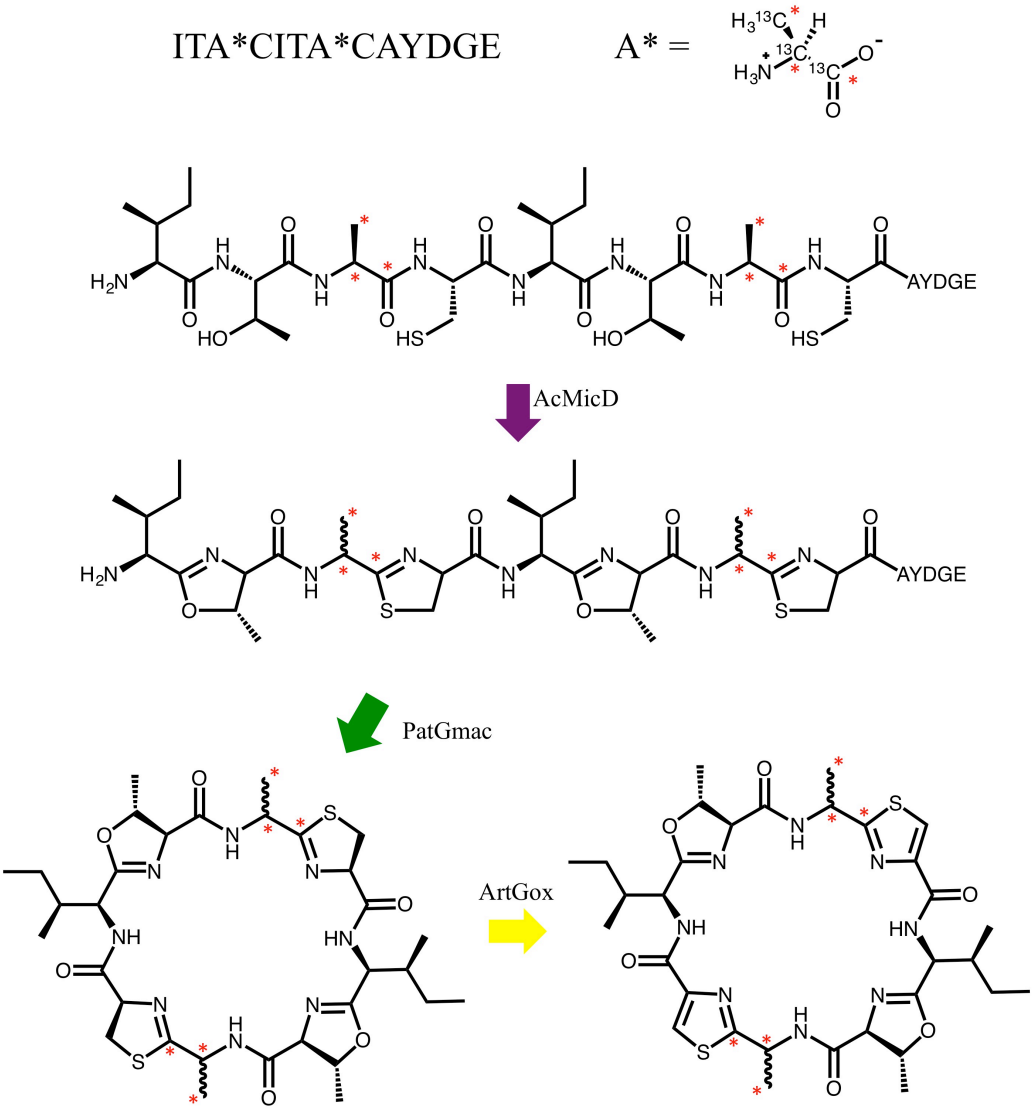



FIGURE 4.2: **Reaction scheme using the starting peptide ITA*CITA*CAYDGE** ITA*CITA*CAYDGE is heterocyclised using AcMicD macrocyclised by PatGmac and oxidised by ArtGox. Red asterisk indicate $^{13}C$. Wavy bonds indicate that epimerisation can occur theoretically at any point following heterocyclisation and so the precise stereochemistry is unknown.

## 4.2 Materials and Methods

### 4.2.1 Expression and Purification of Biosynthetic Enzymes

AcMicD was expressed from pJexpress 411 and purified as described (section 5.2.2). PatGmac was expressed from pEHISTEV and purified as described (section 5.2.2). ArtGox was expressed from pEHISSUMOTEV and purified as described (section 5.2.2). PatG-DUF$_{di.}$ was expressed from pEHISTEV and purified as described (section 2.2.1).

### 4.2.2 Preparation of Deuterated Conditions for NMR Experiments

A 2.5 mM stock solution of synthetic ITA*CITA*CAYDGE (Biosyn) was prepared in $D_2O$. A stock of one-pot reaction buffer/$D_2O$ (Appendix A) was prepared by dissolving all components in $D_2O$. Stock solutions of 100 mM ATP and 1 M $MgCl_2$ were prepared in one-pot reaction buffer/$D_2O$. Biosynthetic enzymes were exchanged into one-pot reaction buffer/$D_2O$ prior to their addition to the reaction mixture.

### 4.2.3 Heterocyclisation of ITA*CITA*CAYDGE

Synthetic ITA*CITA*CAYDGE was incubated with AcMicD as described (section 5.2.4). Heterocyclised products were purified via size exclusion chromatography by passage down a Highload 16/60 Superdex 30 gel filtration column (GE Healthcare).

### 4.2.4 Macrocyclisation of $IT^{MeOx}A*C^{ThH}IT^{MeOx}A*C^{ThH}AYDGE$

Pure $IT^{MeOx}A*C^{ThH}IT^{MeOx}A*C^{ThH}$AYDGE was incubated with PatGmac as described (section 5.2.4).

### 4.2.5 Oxidation of cyclo[$IT^{MeOx}A*C^{ThH}IT^{MeOx}A*C^{ThH}$]

To oxidise thiazolines to thiazoles, 10 μM ArtGox and 200 μM FMN were added to the macrocyclisation reaction mixture, and incubated at 27 °C for 3 h.

### 4.2.6 Incubation of Macrocyclic Products

To investigate spontaneous epimerisation, macrocyclic products were extracted from the one-pot reaction mixtures using an equal volume of d10-nBuOD, evaporated to dryness using a Thermo Savent Speed Vac, resolubilised in $D_2O$ supplemented with 5 % d6-DMSO incubated at 27 °C.

### 4.2.7 $^{1}$H-$^{13}$C HSQC NMR measurements

For NMR experiments, 300 µL samples were transferred to a Shigemi NMR tube (CortecNet) matched to the appropriate solvent ($D_2O$ or d6-DMSO). NMR experiments were performed at 22 °C on a Bruker Ultrashield 700 sprectrometer. The instrument was ran and the data analysed using *TopSpin* software (Bruker).

### 4.2.8 Measuring Macrocyclisation Rate by HPLC

Macrocyclisation rate experiment was performed as described (section 2.4.12).

## 4.3 Results and Discussion

### 4.3.1 Investigating Epimerisation on Linear Peptide

A sample of ITA*CITA*CAYDGE (100 µM) was prepared in one-pot reaction buffer/$D_2O$ (Appendix A) and a $^{1}$H-$^{13}$C HSQC NMR spectrum was recorded (Fig. 4.3).

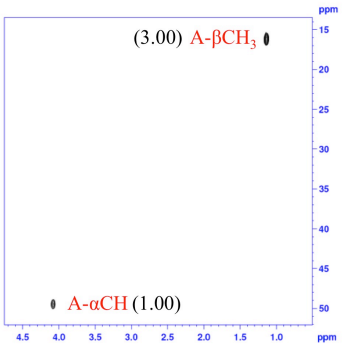
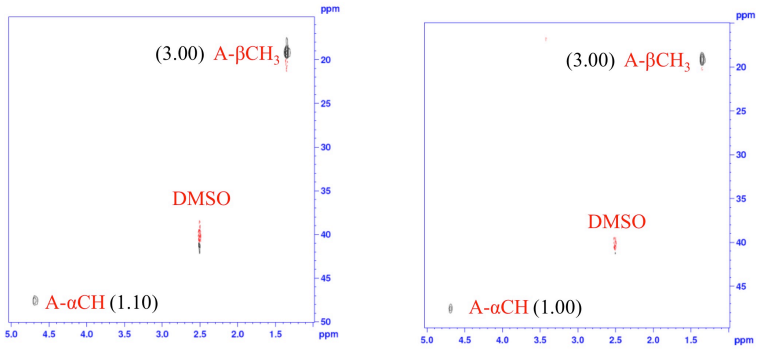



FIGURE 4.3: **ITA*CITA*CAYDGE $^{1}$H-$^{13}$C HSQC NMR spectrum** Spectrum was recorded in one-pot reaction buffer/$D_2O$. Integrals for the αCH and $\beta CH_3$ peaks are shown in brackets

Despite the peptide containing two labelled alanine residues, both the residues are in highly similar chemical environments, such that the signals overlap, and the $^{1}$H-$^{13}$C HSQC NMR spectrum contains only two, easily identifiable cross-peaks: αCH at 4.08, 49.5 ppm and $\beta CH_3$ at 1.11, 16.2 ppm.

Two heterocyclisation reactions of ITA*CITA*CAYDGE were prepared using AcMicD in one-pot reaction buffer, one made up in $H_2O$ and the other in $D_2O$. The heterocyclised products of each reaction were purified via size exclusion chromatography to remove ATP, which is added in excess to the reaction and might complicate the $^{1}H$-$^{13}C$ HSQC NMR spectrum. The pure products, $IT^{MeOx}A*C^{ThH}IT^{MeOx}A*C^{ThH}AYDGE$ were concentrated to 500 µL, dried to a powder under vacuum and resolubilised to 100 µM in d6-DMSO for NMR analysis (Fig. 4.4).




(a) (b)

FIGURE 4.4: **$IT^{MeOx}A*C^{ThH}IT^{MeOx}A*C^{ThH}AYDGE$ $^{1}H$-$^{13}C$ HSQC NMR spectra** from samples prepared in (a) $H_2O$ and (b) $D_2O$. Both spectra contain clear αCH and $\beta CH_3$ cross-peaks. Integrals for the αCH and $\beta CH_3$ peaks are shown in brackets. Spectra were recorded in d6-DMSO

Following the heterocyclisation reaction, the αCH cross-peak shifted slightly upfield in the hydrogen dimension to 4.7 ppm, which is consistant with heterocyclisation of the neighbouring cysteine[51]. The retention of the αCH cross-peak in the $D_2O$ sample indicates that epimerisation was not concurrent with the heterocyclisation reaction. A second batch of pure $IT^{MeOx}A*C^{ThH}IT^{MeOx}A*C^{ThH}AYDGE$, transformed by AcMicD in one-pot buffer/$D_2O$ was prepared and divided into two samples. One sample was incubated at 100 µM in one-pot buffer/$D_2O$ and the other in one-pot buffer/$D_2O$ with the addition of 20 µM PatG-$DUF_{di.}$, a possible epimerase candidate. The samples were incubated at 27 °C for 48 h before recording $^{1}H$-$^{13}C$ HSQC NMR spectra (Fig. 4.5).

The $^{1}H$-$^{13}C$ HSQC NMR spectra of $IT^{MeOx}A*C^{ThH}IT^{MeOx}A*C^{ThH}AYDGE$ contain a clear αCH cross peak following 48 h incubation in both the presence and the absence of PatG-$DUF_{di.}$. Although the integral of the αCH peak relative to the $\beta CH_3$ is lower than before, the same ratio between the integrals is observed for a control reaction with PatG-$DUF_{di.}$ in $H_2O$ (Fig. 4.5), indicating that the loss of the αCH signal is not a result

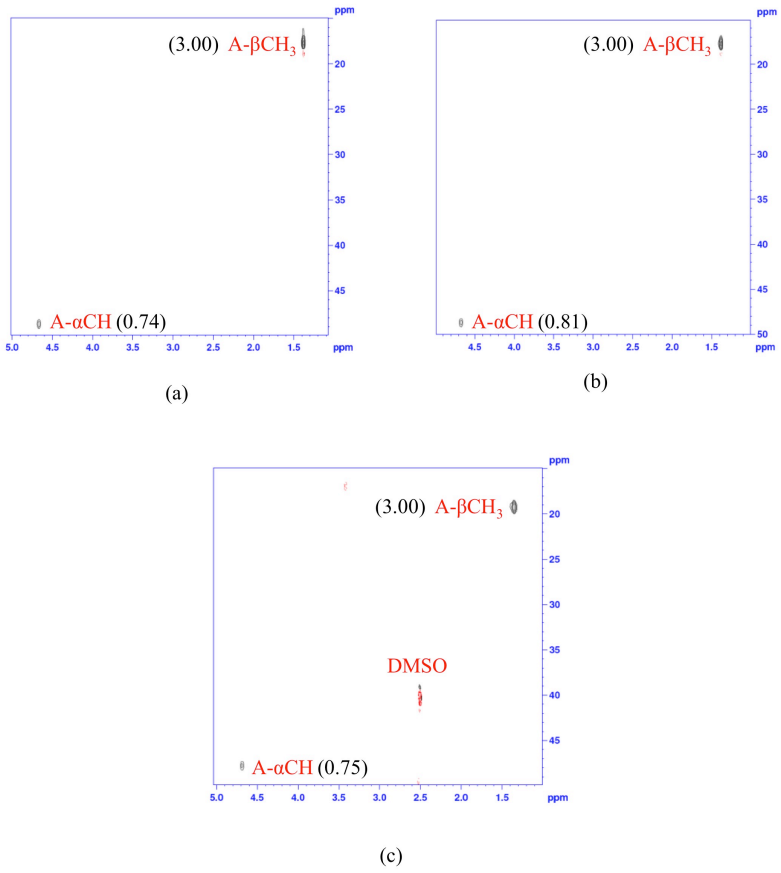



FIGURE 4.5: **$IT^{MeOx}A^*C^{ThH}IT^{MeOx}A^*C^{ThH}$AYDGE incubations $^1$H-$^{13}$C HSQC NMR spectra** (a) one-pot buffer/$D_2O$ and (b) one-pot buffer/$D_2O$ + 20 μM PatG-$DUF_{di.}$. (c) Control sample of PatG-$DUF_{di.}$ incubation in one-pot buffer/$H_2O$. All three spectra contain clear αCH and $\beta CH_3$ cross-peaks. Integrals for the αCH and $\beta CH_3$ peaks are shown in brackets. Spectra were recorded in one-pot reaction buffer/$D_2O$. The control sample was recorded in d6-DMSO.

of H/D exchange. The reason for the loss in signal is unknown. It is assumed the one-pot buffer conditions (pH 9.0) are sufficiently basic to abstract the labile alanine αCH after incubation for two days if epimerisation is spontaneous. Therefore, these data indicate that epimerisation is not spontaneous on the linear peptide following heterocyclisation, and that PatG-$DUF_{di.}$ does not function as an epimerase, at least under the conditions tested.

### 4.3.2 Epimerisation of the Macrocycle

Pure IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$AYDGE was incubated with PatGmac in one-pot reaction buffer/$D_2O$ under standard conditions (section 5.2.4). After 1 week a $^{1}$H-$^{13}$C HSQC NMR spectrum was recorded (Fig. 4.6).

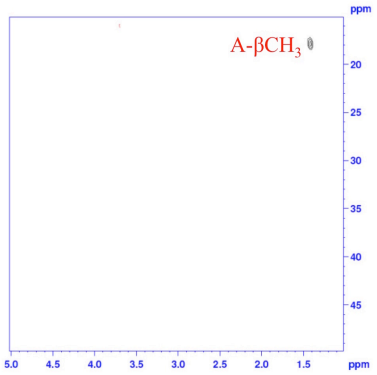
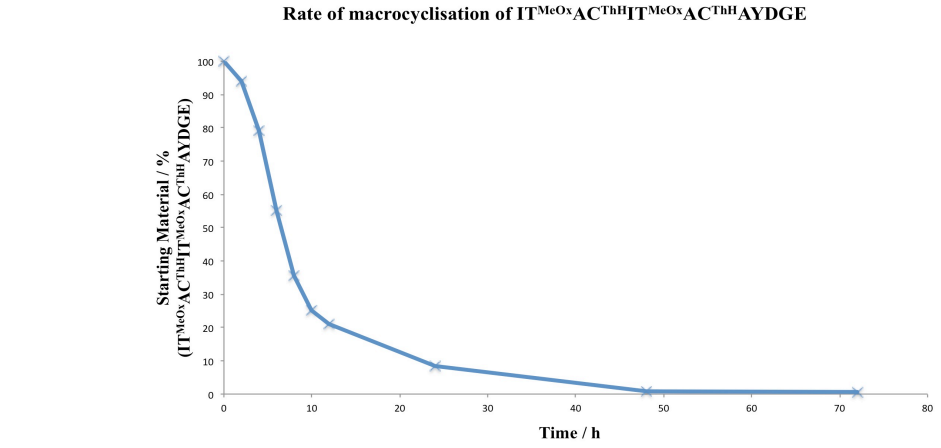


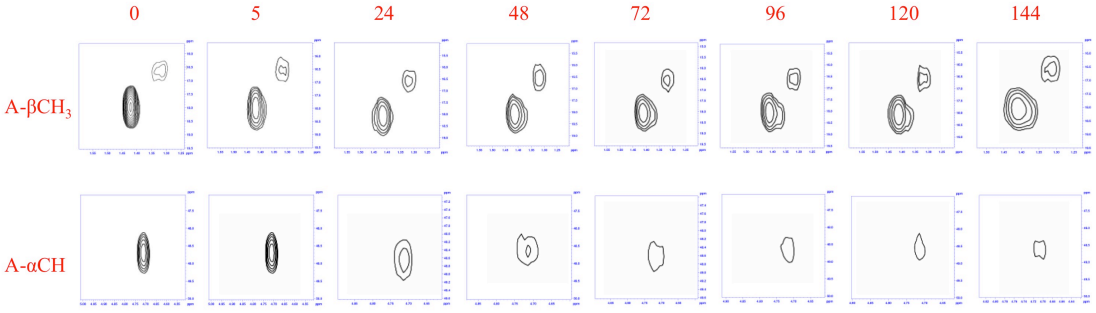

FIGURE 4.6: **IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$AYDGE + PatGmac $^{1}$H-$^{13}$C HSQC NMR spectrum** αCH is no longer visible. Spectrum was recorded in one-pot reaction buffer/$D_2O$.

The absence of the αCH cross-peak, while the βCH$_3$ is still present shows that H/D exchange has occurred at this position, and suggests the stereocentre has epimerised. These data imply that either epimerisation takes place spontaneously once the macrocycle has formed, or that it is someway catalysed by PatGmac, either as a mechanistic consequence of the macrocyclisation reaction, or as a separate reaction. To determine which is correct, it is necessary to understand the relative rates of the macrocyclisation and epimerisation reactions. The rate of macrocyclisation was investigated using HPLC to follow consumption of the unlabelled substrate IT$^{MeOx}$AC$^{ThH}$IT$^{MeOx}$AC$^{ThH}$AYDGE as described (section 2.2.12). From this experiment it is clear that 90 % of the starting material had been consumed after 24 h, and that the reaction was essentially complete (<1 % starting material) by 48 h (Fig. 4.7). A new macrocyclisation reaction of the labeled IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$AYDGE was prepared and monitored by $^{1}$H-$^{13}$C HSQC NMR at 0, 5, 24, 48, 72, 96, 120 and 144 h (Fig. 4.7). Overtime the intensity of the αCH cross-peak is significantly reduced, while the intensity of the βCH$_3$ remains constant. While the macrocyclisation reaction can be considered 90 % complete by 24 h, the intensity of the αCH cross-peak does not decrease as rapidly, and continues to decrease up to 144 h. This observed delay between macrocyclisation and epimerisation

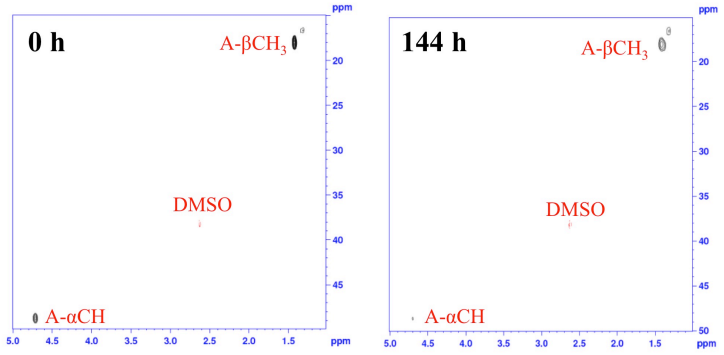
demonstrates that epimerisation is not a mechanistic consequence of macrocyclisation. These data indicate that it either occurs spontaneously once the macrocycle has been formed, or it is catalysed by PatGmac in a completely separate reaction.




(a)




(b)




(c)

**Integral ratio**

| Time /h | αCH | βCH$_3$ |
|---|---|---|
| 0 | 0.74 | 3.00 |
| 5 | 0.65 | 3.00 |
| 24 | 0.54 | 3.00 |
| 48 | 0.50 | 3.00 |
| 72 | 0.33 | 3.00 |
| 96 | 0.29 | 3.00 |
| 120 | 0.34 | 3.00 |
| 144 | 0.22 | 3.00 |

(d)

FIGURE 4.7: **Rate of macrocyclisation vs epimerisation** (a) Consumption of IT$^{MeOx}$AC$^{ThH}$IT$^{MeOx}$AC$^{ThH}$AYDGE substrate during the macrocyclisation reaction over time. The reaction is essentially complete by 48 h. (b) $^{1}$H-$^{13}$C HSQC NMR spectra of IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$AYDGE + PatGmac at various time intervals. A close of the βCH$_3$ and αCH cross-peaks are shown for clarity (c) Full-size $^{1}$H-$^{13}$C HSQC NMR spectra of the first (0 h) and last (144 h) time point. (d) Table showing integrals of αCH and βCH$_3$ peaks at each time point.

To investigate whether PatGmac is responsible for epimerisation, the experiment was repeated, but this time the macrocycle was extracted from the reaction mixture using d10-n-BuOD after 24 h (as described in section 5.3.3) and exchanged back into one-pot reaction buffer in the absence of enzyme. A $^{1}$H-$^{13}$C HSQC NMR spectrum was recorded and signal for the αCH was still clearly visible (Fig. 4.8). The sample was then incubated at 27 °C for a further 24 h and a second $^{1}$H-$^{13}$C HSQC NMR spectrum was recorded. Comparing the two spectra, it is clear that the αCH cross-peak is much smaller following the 24 h incubation (figure 4.7), confirming that H/D exchange is not catalysed by PatGmac, and so it is a spontaneous reaction.

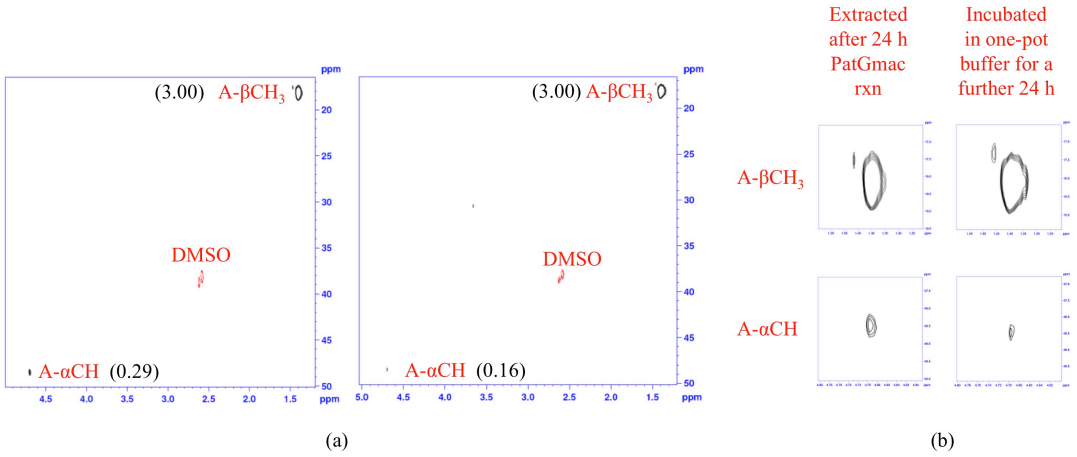



FIGURE 4.8: **Epimerisation of the macrocycle in the absence of PatGmac** (a) $^{1}$H-$^{13}$C HSQC NMR spectra of the extracted cyclo[IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$] (left) and cyclo[IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$] after 24 h incubation in one-pot reaction buffer (right). Integrals for the αCH and βCH$_3$ peaks are shown in brackets (b) Close-ups of the βCH$_3$ and αCH cross-peaks.

Exchange of the αCH signal observed in these experiments is not proof of epimerisation. In nature it is expected that epimerisation of these labile stereocentres results in a more thermodynamically stable conformation of the macrocycle, and thus the epimerisation reaction is unidirectional[67]. Consequently if epimerisation of cyclo[IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$] has occurred then the deuterium which has been incorporated into the cyclic peptide would not be expected to exchange back. To test this, a sample of extracted cyclo[IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$], which had been incubated in one-pot buffer/D$_2$O to allow for H/D exchange, was evaporated to dryness, and resolubilised in one-pot buffer/H$_2$O and incubated at 27 °C. After 1 week the sample was dried and resolubilised in d6-DMSO and a $^{1}$H-$^{13}$C HSQC NMR spectrum was recorded (Fig. 4.9). Despite prolonged incubation in H$_2$O no exchange of molecular

deuterium with solvent hydrogen was observed, indicating that following the observed H/D exchange, the macrocycle exists in a thermodynamically stable conformation, with a high kinetic energy barrier preventing back exchange, suggesting epimerisation has indeed occurred.

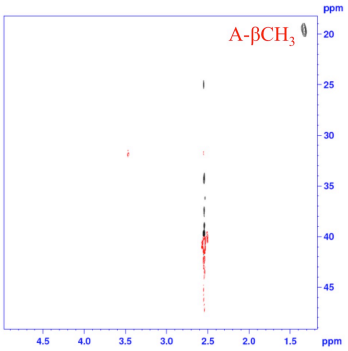



FIGURE 4.9: **Irreversible H/D exchange** $^{1}$H-$^{13}$C HSQC NMR spectra of the extracted cyclo[IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$] following incubation in one-pot buffer/$H_2O$ for 1 week. The αCH cross-peak is not visible indicating H/D exchange does not reoccur under these conditions. The spectrum was recorded in d6-DMSO

.

All the experiments were conducted at pH 9.0. To explore the effect of pH on the epimerisation reaction a sample of extracted macrocycle was resolubilised in one-pot reaction buffer at pH 7.6 and a $^{1}$H-$^{13}$C HSQC NMR spectrum was recorded (Fig. 4.10). As before the αCH is clearly visible. The sample was incubated at 27 °C and further $^{1}$H-$^{13}$C HSQC NMR spectra were recorded after 24 h and 6 d (Fig. 4.10). Although little change was observed after 24 h, after 6 d the αCH signal had completely vanished, suggesting epimerisation is slower at pH 7.6 but not precluded.

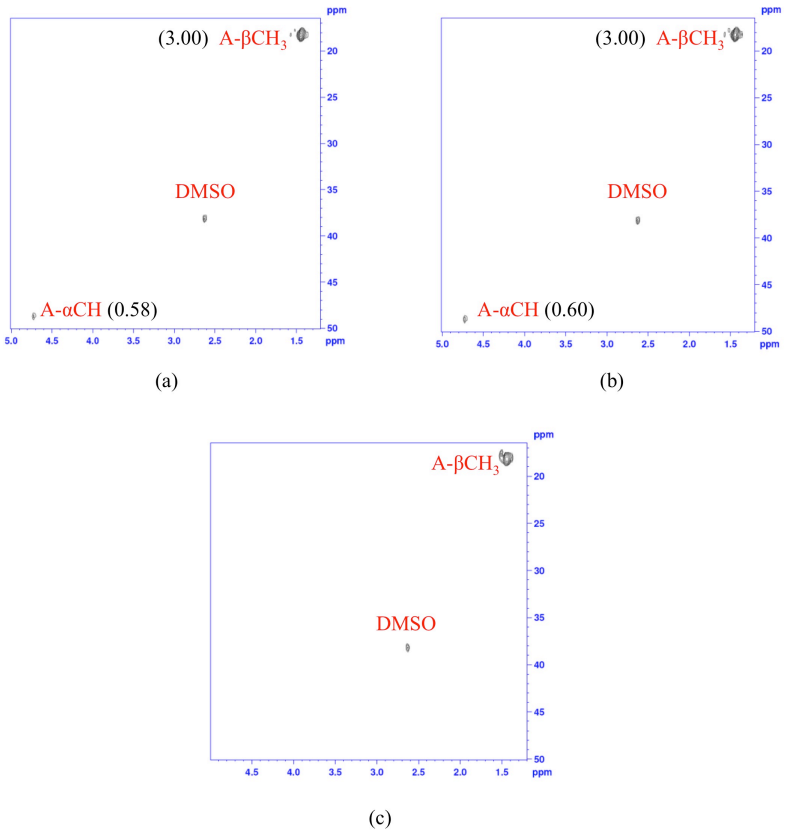



FIGURE 4.10: **Epimerisation at pH 7.6** $^{1}$H-$^{13}$C HSQC NMR spectra of the extracted (a) cyclo[IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$], (b) cyclo[IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$] after 24 h, and (c) 6 d incubation in one-pot reaction buffer/$D_2O$ pH 7.6. Integrals for the αCH and β$CH_3$ peaks are shown in brackets.

It has been proposed that, on the natural compounds, epimerisation occurs after heterocyclisation, but prior to oxidation [45]. If true, then by oxidising the thiazoline heterocycles quickly after macrocyclisation, it might be possible to trap the adjacent residues in the L-conformation. To test this, pure IT$^{MeOx}$A*C$^{ThH}$IT$^{MeOx}$A*C$^{ThH}$AYDGE was incubated with PatGmac in one-pot reaction buffer/$D_2O$ under standard conditions. After 24 h, 10 μM ArtGox and 200 μM FMN were added, and the reaction was incubated for 3 h. The oxidised macrocyclic product, cyclo[IT$^{MeOx}$A*C$^{Thz}$IT$^{MeOx}$A*C$^{Thz}$] was extracted using d10-n-BuOD and exchanged back into one-pot reaction buffer/$D_2O$ and a $^{1}$H-$^{13}$C HSQC NMR spectrum was recorded (Fig. 4.11). The sample was incubated for a further 24 h and a second $^{1}$H-$^{13}$C HSQC NMR spectrum was recorded (Fig. 4.11). Similarly to the reduced compound, the intensity of the αCH cross-peak of cyclo[IT$^{MeOx}$A*C$^{Thz}$IT$^{MeOx}$A*C$^{Thz}$] decreases after 24 h incubation, indicating H/D exchange is still possible on the oxidised compound.

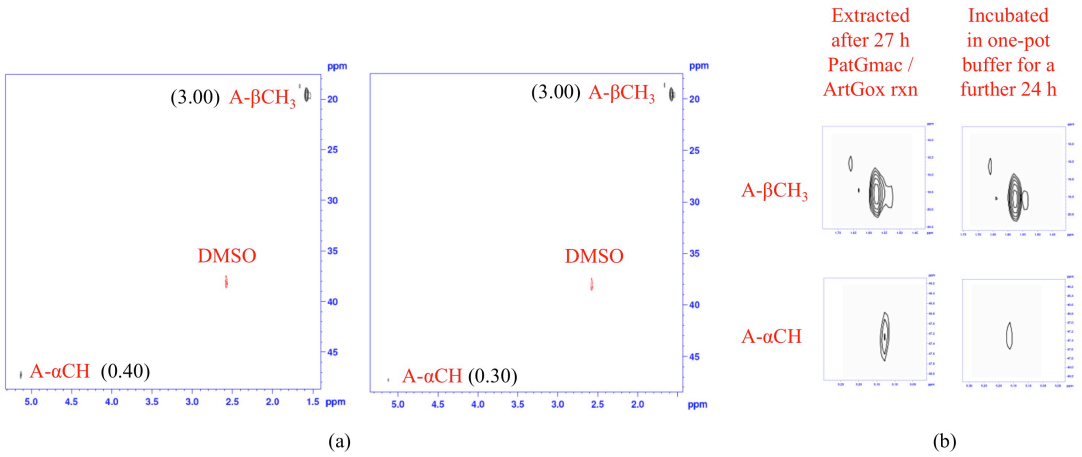



FIGURE 4.11: **Epimerisation of the oxidised macrocycle** (a) $^{1}$H-$^{13}$C HSQC NMR spectra of the extracted cyclo[IT$^{MeOx}$A*C$^{Thz}$IT$^{MeOx}$A*C$^{Thz}$] (left) and cyclo[IT$^{MeOx}$A*C$^{Thz}$IT$^{MeOx}$A*C$^{Thz}$] after 24 h incubation in one-pot reaction buffer/$D_2O$(right). Integrals for the αCH and $\beta CH_3$ peaks are shown in brackets. (b) Close-ups of the $\beta CH_3$ and αCH cross-peaks.

Surprisingly the thermodynamic driving force of the favourable conformation that results from the epimerisation reaction is sufficiently strong, such that epimerisation is spontaneous and inevitable.

## 4.4 Conclusions

The $^{1}$H-$^{13}$C HSQC NMR experiments described in this chapter are the first attempt to definitively address the uncertainties in cyanobactin epimerisation, including when during cyanobactin synthesis epimerisation takes place, and whether it is a spontaneous or enzymatic process. Understanding cyanobactin epimerisation is crucial when attempting to characterise natural products, as well as analogues made in a laboratory. Furthermore, it is important to know whether an epimerisation reaction needs to be built into the one-pot *in vitro* synthesis method (chapter 5).

Although the $^{1}$H-$^{13}$C HSQC NMR experiments technically monitor for exchange of labile αC hydrogen with solvent deuterium, it is assumed this correlates with epimerisation. No H/D exchange, and thus we concluded no epimerisation, was observed on the heterocyclised linear peptide. This fits with the 'conformational stability' hypothesis (section 1.4.5) as it is difficult to expect one conformation (L or D) to be significantly favoured over the other on a flexible linear peptide. PatG-$DUF_{di.}$ was tested as possible

epimerase candidate. Although PatG-DUF$_{\text{di.}}$ had been shown previously not to interact with full length PatE′ or a heterocyclised derivative, an interaction of PatG-DUF$_{\text{di.}}$ with a leaderless peptide substrate had not been ruled out (Chapter 2). Incubation of IT$^{\text{MeOx}}$A*C$^{\text{ThH}}$IT$^{\text{MeOx}}$A*C$^{\text{ThH}}$AYDGE yielded an identical result with and without PatG-DUF$_{\text{di}}$, confirming PatG-DUF$_{\text{di}}$ does not catalyse epimerisation on the linear peptide.

Loss of the αCH cross-peak in $^{1}$H-$^{13}$C HSQC NMR spectra following macrocyclisation indicated epimerisation was either spontaneous on the macrocycle or catalysed by PatGmac; the latter was ruled out when H/D exchange was observed on cyclo[IT$^{\text{MeOx}}$A*C$^{\text{ThH}}$IT$^{\text{MeOx}}$A*C$^{\text{ThH}}$] in the absence of PatGmac at both pH 9.0 and pH 7.6. Furthermore, this exchange was found to be irreversible following incubation of the αC-D containing cyclo[IT$^{\text{MeOx}}$A*C$^{\text{ThH}}$IT$^{\text{MeOx}}$A*C$^{\text{ThH}}$] in buffered $H_2O$. This indicates that the initial H/D exchange did correlate with epimerisation, resulting in an irreversible formation of a thermodynamically favourable conformation, which is consistant with the pre-stated 'conformational stability' hypothesis (section 1.4.5). Interestingly, exchange was still observed following oxidation of thiazoline to thiazole, emphasising the strength of this thermodynamic driving force. These data indicate that patellamide D produced via the one-pot *in vitro* synthesis method (chapter 5) should be authentic, and that an epimerisation reaction should not need to be built into the synthesis. Furthermore the persistence of the epimerisation reaction, for stereocentres adjacent to thiazole moieties as well as thiazoline heterocycles mean that the stereochemistry of the final compounds produced via the one-pot *in vitro* synthesis method is controlled by the conformation of the final product.

## 4.5 Future Work

Under the current one-pot *in vitro* synthesis reaction conditions, it does not seem possible to produce patellamide analogues where the D-amino acids are replaced with L-amino acids. Given the conformational driving force of the reaction, it is possible that factors altering conformation of the cyclic peptide, also affect the propensity for the cyclic peptide to undergo epimerisation. For example, what is the effect of different numbers, and combinations of heterocycles around the macrocycle? What is the effect of *N*-methylation? What is the effect of the epimerisable side-chain? These

experiments were conducted using alanine, the most innocuous chiral amino acid. It seems likely that amino acids with bulkier side-chains would experience more steric pressure, and accordingly a faster rate of epimerisation. What is the effect of the macrocycle ring size - does there become a point where the macrocycle is sufficiently large that residues adjacent to thiazoline/thiazole, depending on their side chain, do not experience significant steric pressure, and so do not epimerise? A number of further experiments are to be conducted in an attempt to answer these questions and ascertain some rules governing the spontaneity of the epimerisation reaction. In addition it should be possible to model these compounds *in silico*, which might provide supporting data for the experiments performed in this chapter, and begin to address remaining uncertainties.

# Chapter 5

# One-pot *in vitro* Synthesis of Diverse Cyclic Peptides

## 5.1 Introduction

The cyanobactin superfamily of highly modified cyclic peptide natural products exhibits a range of interesting bioactive molecules, which could potentially be exploited as therapeutic agents[21]. However, since the first cyanobactins were isolated in the early 1980s[109], their pharmaceutical development has been slow, owing to a low availability of natural material, coupled with technically challenging and expensive syntheses[17]. The enzymes responsible for cyanobactin biosynthesis display low substrate specificity, and are highly tolerant towards substitutions within a precursor peptide's core peptide[27,40,42–44,47,51,59]. It is hypothesised that this promiscuity can be exploited to design novel cyanobactin analogues, enabling the determination of structure-activity relationships. Ultimately bespoke macrocycles, displaying enchanced efficacy when compared to their natural counterparts, can be engineered. Cyanobactin biosynthesis has been reconstituted *in vitro*, making use of various enzymes from different cyanobactin pathways to produce natural and unnatural azol(in)e containing cyclic peptides on a milligram scale[47]. Using the *in vitro* method, engineered PatE precursor peptides, containing various core peptide sequences, can be combined with different cyanobactin processing enzymes, generating multiple products from a single precursor

peptide. To date, this method has been used to produce 17 macrocycles, containing 6-9 residues, representing 11 out of the 20 canonical amino acids[47].

A major drawback of the current *in vitro* synthesis is the use of trypsin. Although trypsin offers a rapid alternative to PatA, the non-specific nature of its proteolysis prevents the use of positively charged amino acids in the core peptide sequence and its use demands multiple purification steps, affecting the yield and the efficiency of the whole process. The ability of AcLynD to process short, leaderless substrates negates this problem. AcMicD and AcPatD (homologues of AcLynD) were cloned (by Ms Ying Ge, PhD student), capable of processing serine and threonine residues in addition to cysteine residues in short leaderless substrates. Using the AcD enzymes an efficient one-pot synthesis method was developed. The ease and efficiency of this method to produce the macrocycle cyclo[IT$^{\mathrm{MeOX}}$AC$^{\mathrm{Thz}}$IT$^{\mathrm{MeOx}}$AC$^{\mathrm{Thz}}$] (**1**) from a synthetic peptide with the sequence ITACITACAYDGE is compared to the original *in vitro* synthesis using a PatE′ variant with the same core peptide sequence (PatE′-ITACITAC). For characterisation of the *in vitro* synthesis products, the previous purification procedures were optimised, including the incorporation of a simple, preliminary two-phase extraction method to separate cyclic peptide products from enzymes, cofactors and linear peptides. Additionally, a robust method for quantitation of purified cyclic peptide products was developed. The one-pot synthesis method is easily scalable and the ability to use short synthetic substrates greatly increases the diversity available through an *in vitro* synthesis, while the use of enzymes keeps the process green and cost-effective. The one-pot *in vitro* synthesis has been employed to generate a library of structurally-diverse compounds, suitable for obtaining structure-activity relationships in collaboration with Dr Andrew Bent, Scottish Enterprises and Ripptide-Pharma.

## 5.2 Materials and Methods

### 5.2.1 Expression and Purification of PatE′-ITACITAC

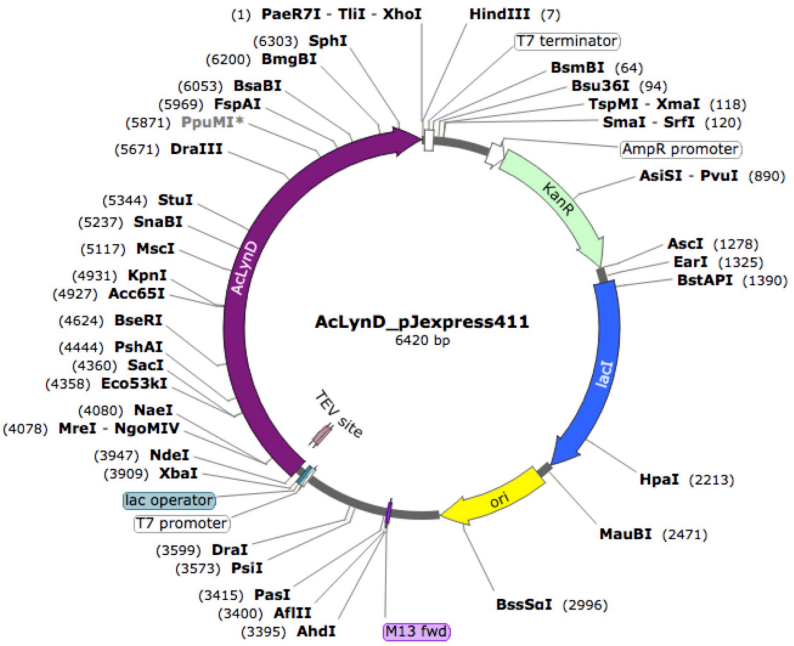
A variant of PatE′ with the core peptide sequence ITACITAC (PatE′-ITACITAC) was cloned into PBMS23CHIS by Dr Andrew Bent. The full PatE′-ITACITAC amino

acid sequence is as follows: MDKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDA-GLEASKITACITACAYDGELEHHHHHH. PatE′-ITACITAC was expressed and purified as described (section 2.5).

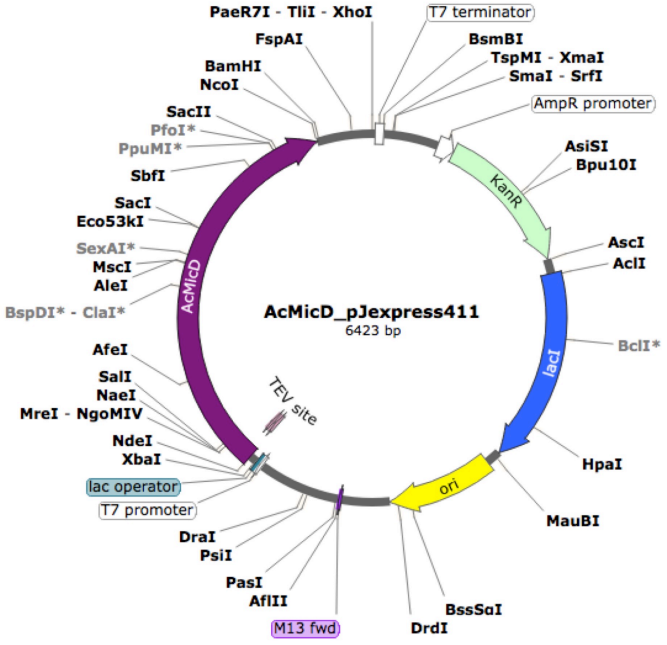
### 5.2.2 Expression and Purification of Biosynthetic Enzymes: LynD, MicD, PatGmac, ArtGox, AcLynD, AcMicD and AcPatD

LynD was expressed from the pJexpress 411 plasmid (DNA 2.0) as described (section 3.2.3). MicD was expressed from the pJexpress 411 plasmid (DNA 2.0) as for LynD. PatGmac was cloned into the pEHISTEV vector (a kind gift from Dr Huanting Liu) by Dr Jesko Koehnke and was expressed as described for LynD. ArtGox was cloned into the pEHISSUMOTEV (a kind gift from Dr David Owen) by Dr Andrew Bent and expressed with a SUMO tag as described for LynD, however the auto-induction media was supplemented with 50 μM riboflavin. AcLynD was cloned and expressed as described (section 3.2.3) AcMicD, AcPatD were cloned into pJexpress 411 (DNA 2.0) by Ms Ying Ge and expressed as described for LynD. Vector maps for AcLynD, AcMicD, AcPatD, PatGmac and ArtGox can be seen in Fig. 5.1.

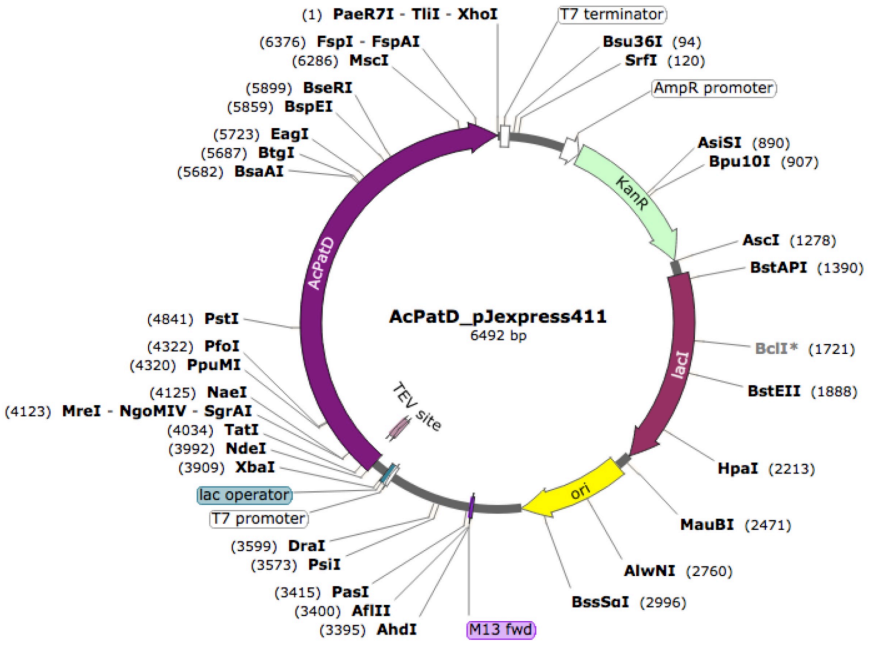
All enzymes were purified as follows: cell pellets were resuspended in general lysis buffer (Appendix A) supplemented with 0.4 mg DNAse (Sigma) per gram of wet cell pellets, and cOmplete EDTA-free protease inhibitor tablets (Roche; 1 per 50 mL resuspension). The cells were lysed via passage through a cell disruptor at 207 MPa (Constant Systems) and clarified by centrifugation (40 000$g$, 4 °C, 20 min). Cleared lysate was passed through a Ni-Sepharose 6 Fast Flow column (GE Healthcare) equilibrated in general lysis buffer. The bound protein was washed with lysis buffer and eluted in general elution buffer (Appendix A). The eluted protein was dialysed (3 x 300 mL, 4 °C) into one-pot storage buffer (Appendix A), concentrated to 1 mM and stored as 1 mL alliquots. For ArtGox general lysis and elution buffers were supplemented with 50 μM FMN; the SUMO tag was not removed during purification.


(1) PaeR7I - TliI - XhoI
HindIII (7)
T7 terminator
(6303) SphI
(6200) BmgBI
BsmBI (64)
Bsu36I (94)
(6053) BsaBI
(5969) FspAI
TspMI - XmaI (118)
SmaI - SrfI (120)
(5871) PpuMI*
AmpR promoter
(5671) DraIII
AsiSI - PvuI (890)
KanR
(5344) StuI
(5237) SnaBI
AscI (1278)
EarI (1325)
(5117) MscI
BstAPI (1390)
(4931) KpnI
(4927) Acc65I
AcLynD
AcLynD_pJexpress411
6420 bp
(4624) BseRI
(4444) PshAI
(4360) SacI
(4358) Eco53kI
lacI
(4080) NaeI
(4078) MreI - NgoMIV
TEV site
(3947) NdeI
(3909) XbaI
HpaI (2213)
lac operator
ori
T7 promoter
MauBI (2471)
(3599) DraI
(3573) PsiI
BssSαI (2996)
(3415) PasI
(3400) AflII
(3395) AhdI
M13 fwd

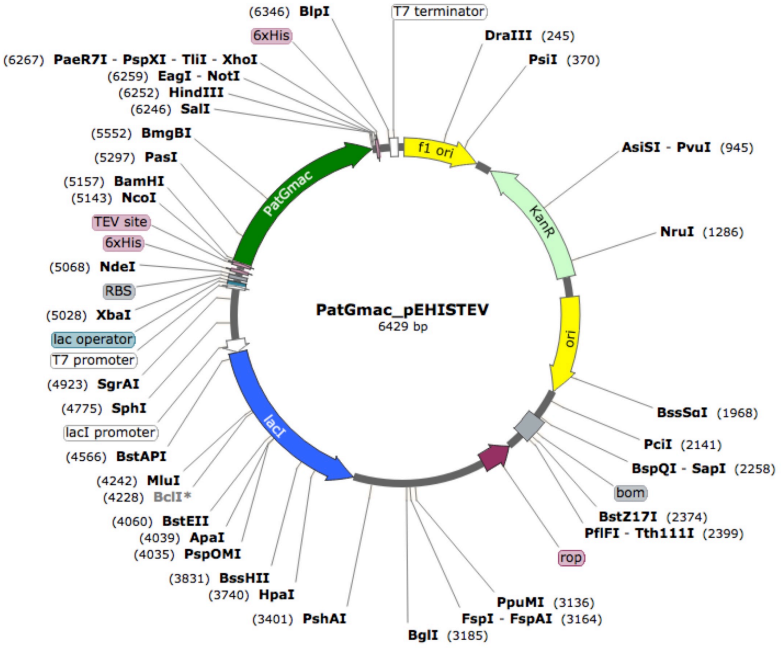
(a)


PaeR7I - TliI - XhoI
T7 terminator
FspAI
BsmBI
TspMI - XmaI
BamHI
NcoI
SmaI - SrfI
AmpR promoter
SacII
PfoI*
AsiSI
PpuMI*
Bpu10I
SbfI
KanR
SacI
Eco53kI
AscI
SexAI*
AclI
MscI
AleI
AcMicD
AcMicD_pJexpress411
6423 bp
BspDI* - ClaI*
BclI*
AfeI
lacI
SalI
NaeI
TEV site
MreI - NgoMIV
HpaI
NdeI
XbaI
ori
lac operator
MauBI
T7 promoter
DraI
PsiI
BssSαI
PasI
AflII
DrdI
M13 fwd

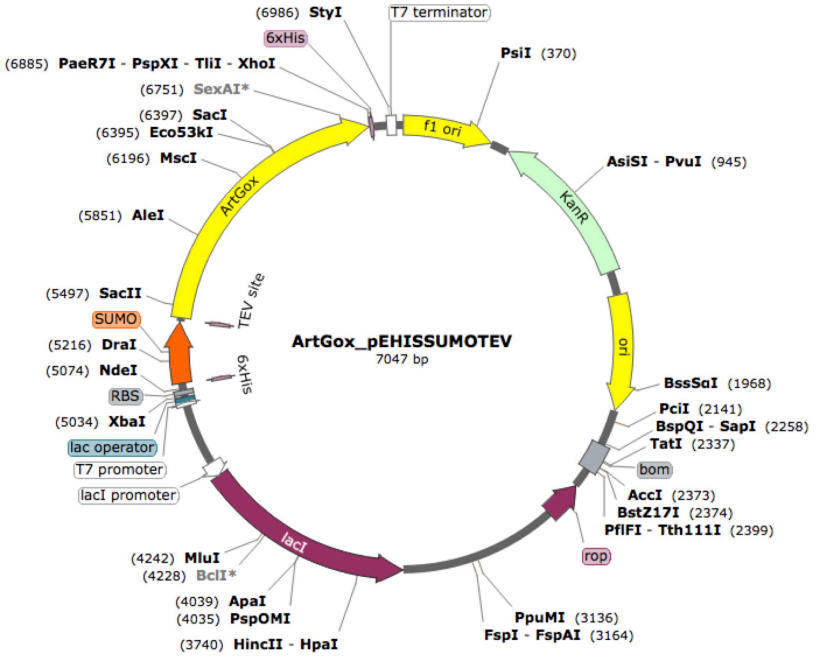
(b)


(1) PaeR7I - TliI - XhoI
T7 terminator
(6376) FspI - FspAI
(6286) MscI
Bsu36I (94)
SrfI (120)
(5899) BseRI
(5859) BspEI
AmpR promoter
(5723) EagI
(5687) BtgI
(5682) BsaAI
AsiSI (890)
Bpu10I (907)
KanR
AscI (1278)
BstAPI (1390)
AcPatD
AcPatD_pJexpress411
6492 bp
(4841) PstI
(4322) PfoI
(4320) PpuMI
BclI* (1721)
lacI
BstEII (1888)
(4125) NaeI
(4123) MreI - NgoMIV - SgrAI
TEV site
(4034) TatI
(3992) NdeI
(3909) XbaI
HpaI (2213)
ori
lac operator
T7 promoter
MauBI (2471)
(3599) DraI
(3573) PsiI
AlwNI (2760)
(3415) PasI
(3400) AflII
BssSαI (2996)
(3395) AhdI
M13 fwd

(c)


(6346) BlpI
T7 terminator
6xHis
DraIII (245)
(6267) PaeR7I - PspXI - TliI - XhoI
PsiI (370)
(6259) EagI - NotI
(6252) HindIII
(6246) SalI
(5552) BmgBI
f1 ori
AsiSI - PvuI (945)
(5297) PasI
(5157) BamHI
(5143) NcoI
PatGmac
KanR
NruI (1286)
TEV site
6xHis
(5068) NdeI
RBS
PatGmac_pEHISTEV
6429 bp
(5028) XbaI
ori
lac operator
T7 promoter
BssSαI (1968)
(4923) SgrAI
(4775) SphI
PciI (2141)
lacI promoter
BspQI - SapI (2258)
(4566) BstAPI
bom
(4242) MluI
(4228) BclI*
lacI
BstZ17I (2374)
(4060) BstEII
PflFI - Tth111I (2399)
(4039) ApaI
(4035) PspOMI
rop
(3831) BssHII
(3740) HpaI
PpuMI (3136)
(3401) PshAI
FspI - FspAI (3164)
BglI (3185)

(d)




(e)

FIGURE 5.1: **Biosynthetic enzyme vectors** (a) AcLynD, (b) AcMicD, (c) AcPatD, (d) PatGmac, (e) ArtGox.

### 5.2.3 *in vitro* Synthesis of cyclo[IT$^{\mathbf{MeOX}}$AC$^{\mathbf{Thz}}$IT$^{\mathbf{MeOx}}$AC$^{\mathbf{Thz}}$] (1) from PatE′-ITACITAC

#### 5.2.3.1 Heterocyclisation of PatE′-ITACITAC using MicD

Pure PatE′-ITACITAC was incubated in the presence of MicD (5 µM), ATP (5 mM) and $MgCl_2$ (5 mM) in gel filtration buffer at 37 °C overnight. The reaction was monitored by MALDI-TOF-MS; a loss of 18 Da, corresponding to the loss of water, was observed for each heterocycle formed. Once the reaction was confirmed to be complete, the product PatE′-IT$^{\mathrm{MeOX}}$AC$^{\mathrm{ThH}}$IT$^{\mathrm{MeOx}}$AC$^{\mathrm{ThH}}$ (**2**) was isolated via size exclusion chromatography by passage down a Highload 16/60 Superdex 75 gel filtration column (GE Healthcare) equilibrated in gel filtration buffer. Pure **2** was concentrated to 10 mg mL$^{-1}$.

#### 5.2.3.2 Removal of Leader Peptide using Trypsin

Pure **2** was incubated with 1:100 trypsin (Sigma-aldrich) at 37 °C, 300 rpm for 4 h. Following incubation, cleaved peptide was isolated using a Ni-Sepharose 6 Fast Flow column (GE Healthcare), and eluted using gel filtration buffer supplemented with 250 mM imidazole. The eluted fraction was passed down a Highload 16/60 Superdex 30 gel filtration column (GE Healthcare), pre-equilibrated in peptide column buffer (Appendix A) to separate cleaved $IT^{MeOX}AC^{ThH}IT^{MeOx}AC^{ThH}AYDGELEH_6$ (**3**) from non-cleaved **2** and to remove imidazole. Purified **3** was concentrated to 300 μM. Successful proteolysis was confirmed using MALDI-TOF-MS.

#### 5.2.3.3 Macrocyclisation with PatGmac

Pure **3** (100 μM) was incubated with PatGmac (20 μM) in peptide column buffer supplemented with DMSO (5 %) and NaCl (350 mM), at 37 °C for 1 week. Reaction progress was monitored using MALDI-TOF-MS.

#### 5.2.3.4 Oxidation of cyclo[$\mathbf{IT^{MeOX}AC^{ThH}IT^{MeOx}AC^{ThH}}$] (4) with ArtGox

Once macrocyclisation was complete, ArtGox (20 μM) and FMN (200 μM) were added to the reaction mixture and the reaction was incubated at 37 °C for 16 h to give **1**.

### 5.2.4 One-pot Synthesis of 1 from Synthetic ITACITACAYDGE (BioSyn)

Synthetic ITACITACAYDGE (110 μM) was heterocyclised by incubation with AcMicD (30 μM), ATP (10 mM), $MgCl_2$ (10 mM) and DTT (0.5 mM) at 27 °C for 16 h in one-pot reaction buffer (Appendix A). Complete heterocyclisation was confirmed by MALDI-TOF-MS. Once complete the modified peptide was diluted to a reacting concentration of 90 μM as PatGmac (50 μM), DMSO (5 %) and NaCl (350 mM) were added to the reaction mixture. The reaction mixture was incubated at 27 °C for 48 h, and macrocyclisation was monitored using MALDI-TOF-MS. To oxidise thiazolines to thiazoles, ArtGox (5 μM) and FMN (200 μM) was added and the reaction mixture was incubated at 27 °C for a further 4 h to give **1**. To oxidise both thiazolines and oxazolines

to thiazoles and oxazoles, 40 μM ArtGox and 1 mM FMN were added to the reaction mixture, and the reaction was incubated at 27 °C for 48 h to give **12**. To produce only thiazoline/thiazole heterocycles the substrate was incubated with AcLynD (5 μM), ATP (5 mM), $MgCl_2$ (5 mM) and DTT (0.5 mM) at 27 °C for 16 h in one-pot reaction buffer. Macrocyclisation and oxidation reactions were performed as described. All peptides in Table 4.2 were processed as described.

### 5.2.5 Extraction of Macrocycles using n-BuOH

Extraction of the macrocyclic product from the reaction mixture was achieved by adding an equal volume of n-butanol (n-BuOH) to the reaction mixture. The two phases were separated by centrifugation (4000*g*, 10 min). The organic phase was collected and the aqueous phase was extracted a further two times with equal volumes of n-BuOH. Using MALDI-TOF-MS, macrocycle was always identified the organic fraction from the first extraction, and sometimes from the second and third extractions. After three extractions, macrocyclic products were never identified in the aqueous phase. The n-BuOH fractions that contained product were evaporated to dryness using a rotary evaporator, and the crude compound was resolubilised in a 1:1 mixture of methanol and water for HPLC analysis and purification.

### 5.2.6 HPLC Purification of Macrocycles

Analytical RP-HPLC was performed on an Agilent infinity 1260 series equipped with a MWD detector and a single quadropole MS using a Nucleodur $C_{18}$ (Macherey-Nagel) column (10 μm x 4.6 x 250 mm). Semi-preparative RP-HPLC was performed on an Agilent infinity 1260 series equipped with a MWD using a Nucleodur $C_{18}$ (Macherey-Nagel) column (10 μm x 16 x 250 mm). Samples were eluted from the column using an aqueous 5 mM ammonium carbonate - MeCN gradient with a flow rate of 1 mL $min^{-1}$. For cyclic peptide products containing two heterocycles (thiazoline/thiazole/proline) or less, the following gradient was used: 0 - 5 min 5 % MeCN, 5 - 40 min 5 - 95 % MeCN, 40 - 45 min 95 % MeCN, 45 - 48 min 95 - 5 % MeCN. For cyclic peptides containing four or three heterocycles (thiazoline/thiazole/oxazoline/oxazole/proline), the following gradient was used: 0 - 5 min 5% MeCN, 5 - 8 min 5 - 50 % MeCN, 8 - 55 min 50 - 95 % MeCN, 55 - 60 min 95 % MeCN, 60 - 62 min 95 - 90 % MeCN. The retention time

of compounds were identified by MS and UV detection at 220 nm using the analytical system. Fractions containing pure cyclic peptide were collected manually using UV detection at 220 nm using the semi-preparative system.

### 5.2.7 NMR Quantitation

All qNMR experiments were performed at 25 °C on a Bruker AVANCE III 500 MHz spectrometer equipped with a room temperature BBFO+ probe. Data acquisition was done with 16 scans, 10.00 compensate 90 ° pulse and a 30 s delay. The instrument was run and the data was analysed using *TopSpin* software (Bruker).

### 5.2.8 Pgp-Glo™ Assay

The Pgp-Glo™ assay detects the effect of compounds on the ATPase activity of recombinant human Pgp in a membrane fraction, by utilising the light-generating ATP-dependant reaction of firefly luciferase (Fig. 5.2). ATP is incubated with Pgp; after a defined time the reaction is stopped and luciferase and luciferin are added to the reaction mixture. The luciferase consumes any remaining unmetabolised ATP to produce oxyluciferin, resulting in a detectable luminescence signal. Compounds that stimulate Pgp ATPase activity, *i.e.* substrates of Pgp transport, result in an increased consumption of ATP by Pgp, and thus a reduction in oxyluciferin production and the concomitant luminescence signal. Conversely, those compounds that inhibit ATPase activity give rise to an increased luminescence signal.

The Pgp-Glo™ assay supplies two control compounds: sodium orthovanadate ($Na_3VO_4$) and verapamil. $Na_3VO_4$ is a selective inhibitor of Pgp, and completely abolishes ATP activity[110]. Any ATPase activity in the presence of $Na_3VO_4$ is attributed to minor non-Pgp contaminants in the membrane fraction. The difference in luminescence signal between the $Na_3VO_4$ treated and untreated Pgp determines the basal Pgp ATPase activity. Comparison of the basal ATPase activity with the activity determined following treatment of Pgp with a test compound, allows compounds to be identified as stimulatory or inhibitory. Verapamil is supplied as a known substrate of, and thus a stimulator of Pgp ATPase activity[111]. By treating Pgp with a test compound in conjunction with

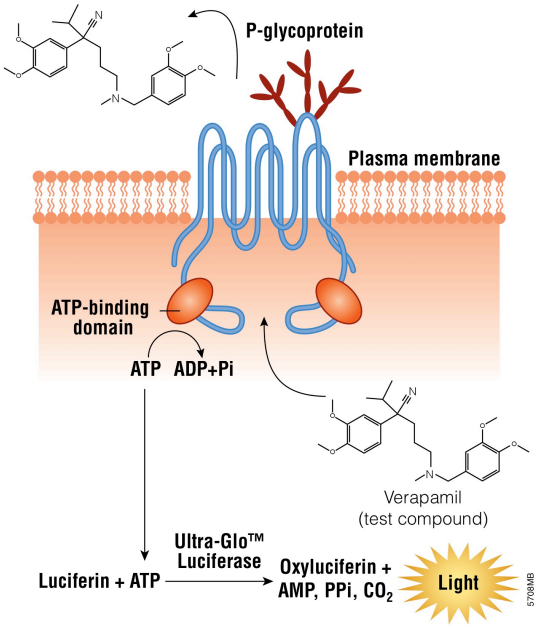



FIGURE 5.2: **Pgp-Glo$^{TM}$ assay** The image was taken from the Pgp-Glo$^{TM}$ Assay technical bulletin.

verpamil, the ability of a test compound to inhibit verapamil stimulated Pgp ATPase activity can be investigated.

Screening and dose-response curves were performed according to manufacturer's instructions, however a longer ATP incubation time of 3 h was used.

## 5.3 Results and Discussion

### 5.3.1 *in vitro* Synthesis of cyclo[IT$^{MeOX}$AC$^{Thz}$IT$^{MeOx}$AC$^{Thz}$] (1)

Full-length PatE′-ITACITAC was expressed and purified as described (section 2.5) yielding 50 mg $L^{-1}$ culture. Purity was assessed by SDS-PAGE and identity confirmed by MALDI-TOF-MS (Fig. 5.3).

Pure PatE′-ITACITAC was heterocyclised with MicD as described (section 5.2.3). The reaction was monitored by MALDI-TOF-MS and the product was purified by size-exclusion chromatography. A mass decrease of 72 Da (m/z = 6842) relative to the

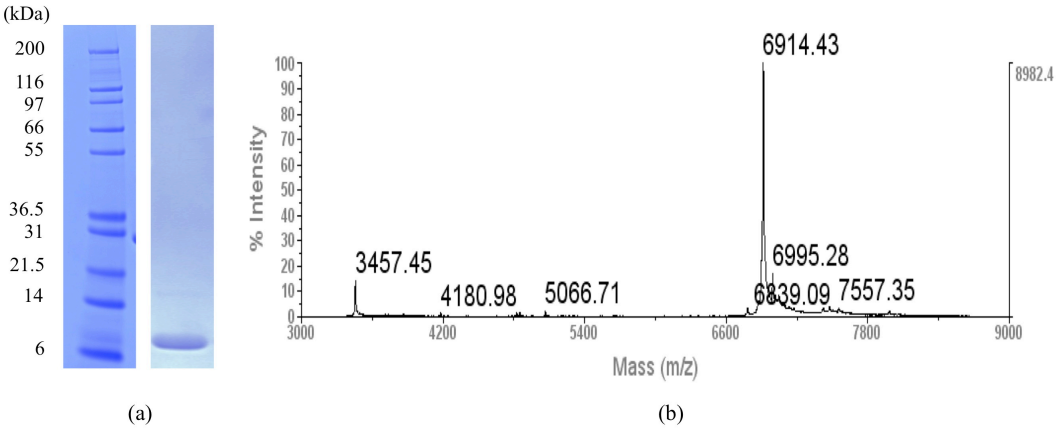



FIGURE 5.3: **PatE′-ITACITAC purification** (a) SDS-PAGE analysis of purified PatE′-ITACITAC (b) MALDI-TOF-MS of purified PatE′-ITACITAC (6913.4 Da).

starting material indicates both threonine and cysteine residues have been heterocyclised giving PatE′-IT$^{\mathrm{MeOX}}$AC$^{\mathrm{ThH}}$IT$^{\mathrm{MeOx}}$AC$^{\mathrm{ThH}}$ (**2**) (Fig. 5.4).

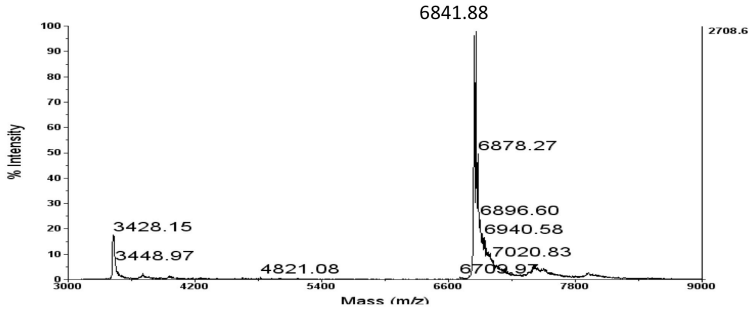



FIGURE 5.4: **PatE′-ITACITAC + MicD** MALDI-TOF-MS of PatE′-ITACITAC + MicD reaction mixture following 16 h incubation at 27 °C.

To remove the leader sequence, **2** was incubated with trypsin as described (section 5.2.4). After 4 h the proteolysis reaction was purified via size exclusion chromatography to separate IT$^{\mathrm{MeOX}}$AC$^{\mathrm{ThH}}$IT$^{\mathrm{MeOx}}$AC$^{\mathrm{ThH}}$AYDGELEH$_6$ (**3**) from trypsin and full length **2**. The presence of **3** was confirmed by MALDI-TOF-MS (Fig. 5.5). However MS analysis of purified **3** (2322.0 Da) revealed a prominent contamination peak at m/z = 1362.8 (Fig. 5.5). Trypsin can cleave C-terminal to any lysine or arginine in a peptide sequence, and so it was expected that the contamination at m/z = 1362.8 corresponds to a fragment of the leader peptide NILPQQGQPVIR (1361.8 Da). Therefore a Ni affinity chromatography step was incorporated into the purification procedure prior to size

exclusion step and was found to sufficiently purify **3** from the undesired NILPQQPVIR peptide (Fig. 5.5).

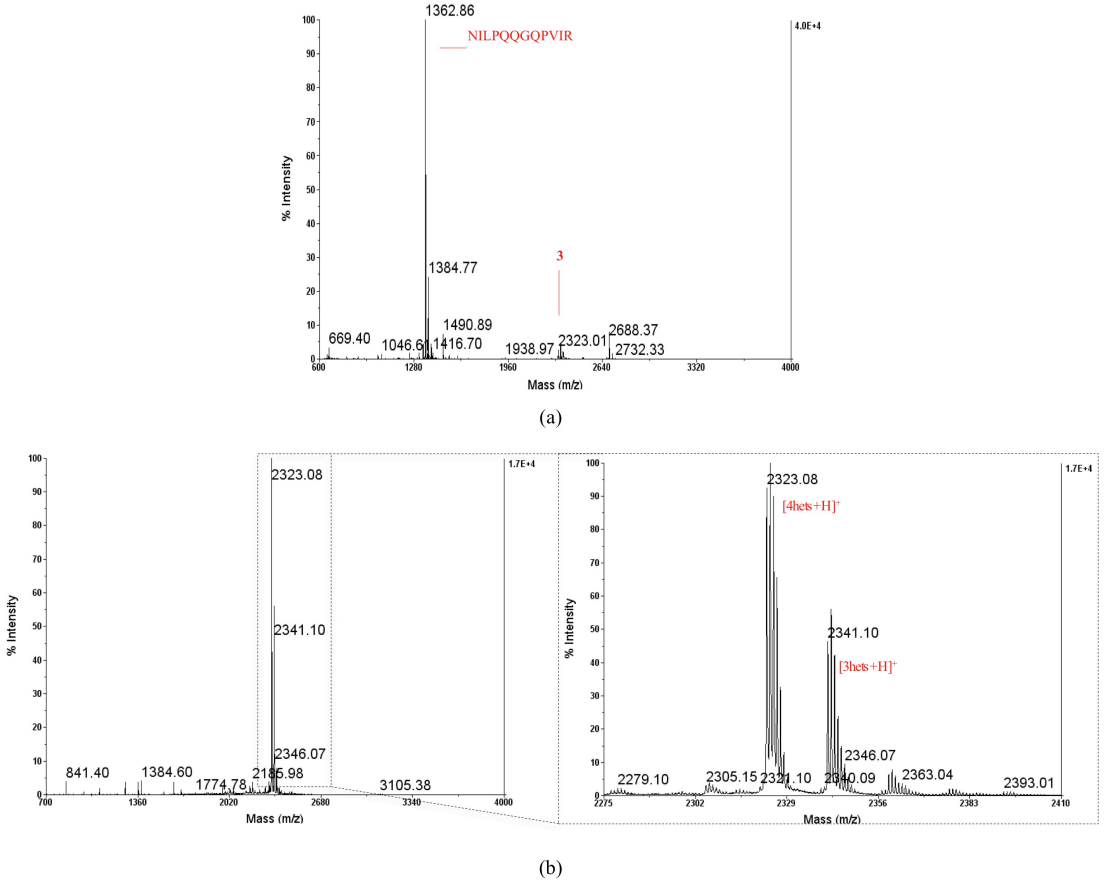



FIGURE 5.5: **2 + trypsin** (a) MALDI-TOF-MS of **3** (2322.0 Da) purified via size exclusion chromatography. A significant contamination with m/z = 1362.8 corresponds tocorrespond to leader peptide fragment NILPQQGPVIR (1361.8 Da). (b) MALDI-TOF-MS of **3** purified by Ni affinity chromatography, showing the contaminating leader peptide fragment has been removed. The dashed box shows an enhanced view of the product peaks.

Pure **3** was incubated with PatGmac as described (section 5.2.5). The reaction was analysed by MALDI-TOF-MS after three days. (Fig. 5.6).

The presence of cyclo[IT$^{\mathrm{MeOX}}$AC$^{\mathrm{ThH}}$IT$^{\mathrm{MeOx}}$AC$^{\mathrm{ThH}}$] (**4**) (m/z 705.4) coupled with the reduction in the intensity of the peak corresponding to **3** (m/z = 2323.0), confirmed the macrocyclisation reaction had progressed. The large peak at m/z = 1618.7 corresponds to the expected side product of the reaction: AYDGELEH$_6$ (1617.7 Da). The adjacent peak at m/z = 1547.6 corresponds to the loss of the N-terminal alanine residue from the side product, but the reason for this is unknown. The reaction was incubated at 37 °C until no starting material was observed by MALDI-TOF-MS (typically 1 week).

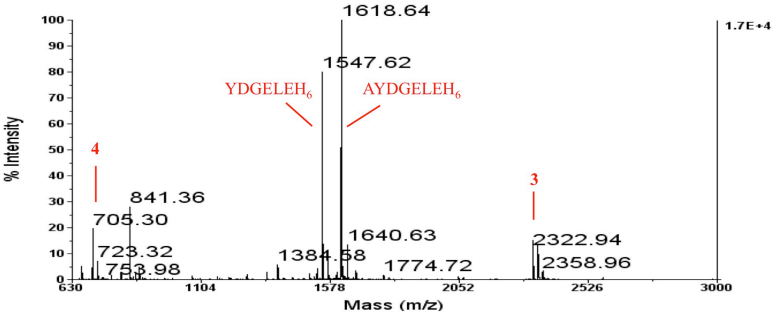
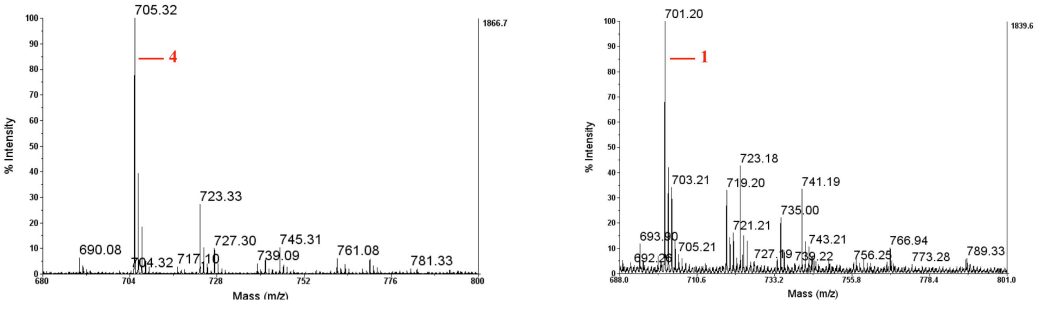



FIGURE 5.6: **3 + PatGmac** MALDI-TOF-MS of **3** + PatGmac. Starting material (2322.0 Da) and **4** (704.4 Da) can be seen. The large peaks at m/z = 1618.7 and 1547.6 correspond to C-terminal fragments $AYDGELEH_6$ and $YDGELEH_6$ respectively.

To avoid a time consuming HPLC purification of the macrocycle, the oxidation reaction was achieved by adding ArtGox and FMN directly to the macrocyclisation reaction mixture. In nature the cyanobactin oxidase and macrocyclase domains are usually part of the same protein and so it was expected each domain would tolerate the same reaction conditions *in vitro*. The oxidation reaction was left overnight and analysed by MALDI-TOF-MS. Oxidation of both thiazolines to thiazoles was identified via a loss of 4 Da (Fig. 5.7).




(a) (b)

FIGURE 5.7: **Oxidation of 4** MALDI-TOF-MS of (a) **4** and (b) **1** following incubation of ArtGox.

### 5.3.2 Development of a One-pot *in vitro* Synthesis of Cyanobactins

Synthetic ITACITACAYDGE (BioSyn) was initially incubated with AcLynD and AcMicD in gel filtration buffer (Appendix A) as described for LynD (section 3.2.4) to form ITAC$^{\text{ThH}}$ITAC$^{\text{ThH}}$AYDGE (**5**) and IT$^{\text{MeOX}}$AC$^{\text{ThH}}$IT$^{\text{MeOx}}$AC$^{\text{ThH}}$AYDGE (**6**) respectively. Under these reaction conditions, AcLynD completed the cyclodehydration of both cysteine residues to thiazolines as expected (Fig. 5.8). However the use of AcMicD resulted in a mixture of products containing 4, 3, and 2 heterocycles overnight (Fig. 5.8). Following iterative rounds of optimisation of the reaction conditions (data not shown), the reaction buffer was changed from gel filtration buffer to one-pot reaction buffer (Appendix A) and the reaction conditions altered to those described (section 5.2.4). These optimised conditions resulted in more efficient processing by AcMicD, such that the major compound was **6** - a species containing 4 heterocyles following overnight incubation (Fig. 5.8). However a large amount of AcMicD is needed to achieve the transformation (30 μM), compared to AcLynD (5 μM). Work in the lab is ongoing to further optimise the heterocyclisation reaction of leaderless peptide substrates, and is not within the scope of this chapter.

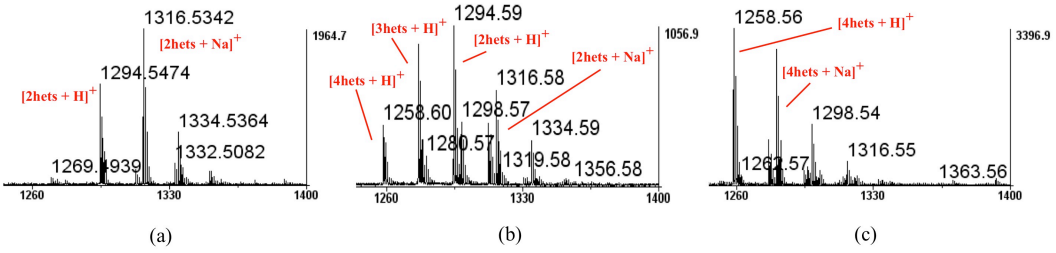



FIGURE 5.8: **ITACITACAYDGE + AcLynD/AcMicD** MALDI-TOF-MS of (a) ITACITACAYDGE + AcLynD under standard reaction conditions in gel filtration buffer gives predominantly a product with 2 heterocycles (**5**) (b) ITACITACAYDGE + AcMicD under standard reaction conditions in gel filtration buffer gives a mixture of compounds with 4, 3 and 2 heterocycles (c) ITACITACAYDGE + AcMicD under new reaction conditions in one-pot reaction buffer gives predominantly the desired product with 4 heterocycles (**6**).

Following the formation of **6**, there is a choice as to which enzymatic reaction, either macrocyclisation or oxidation, should be incorporated as the second step in a one-pot *in vitro* synthesis of **1**. Little is known about the relative order of the macrocyclisation and oxidation reactions in nature, with both hypothetically possible after heterocyclisation and neither necessarily, a prerequisite of the other[28]. However, the cyanobactin

macrocyclisation reaction is very slow on thiazoline-containing peptides *in vitro* [43,55] and it is conceivable that oxidation of thiazolines to thiazoles on the linear peptide, provides a more optimal substrate for the macrocyclase, resulting in a favourable rate acceleration of macrocyclisation. To test this hypothesis, three one-pot synthesis strategies were devised: (1) sequential addition of AcD, oxidase and macrocyclase; (2) simultaneous addition of AcD and oxidase, followed by addition of the macrocyclase; (3) sequential addition of AcD, macrocyclase and oxidase (Fig. 5.9) Of these strategies, strategy (3) follows the same enzymatic order of the previous *in vitro* synthesis, and so is expected to work, thus acting as a control [47].

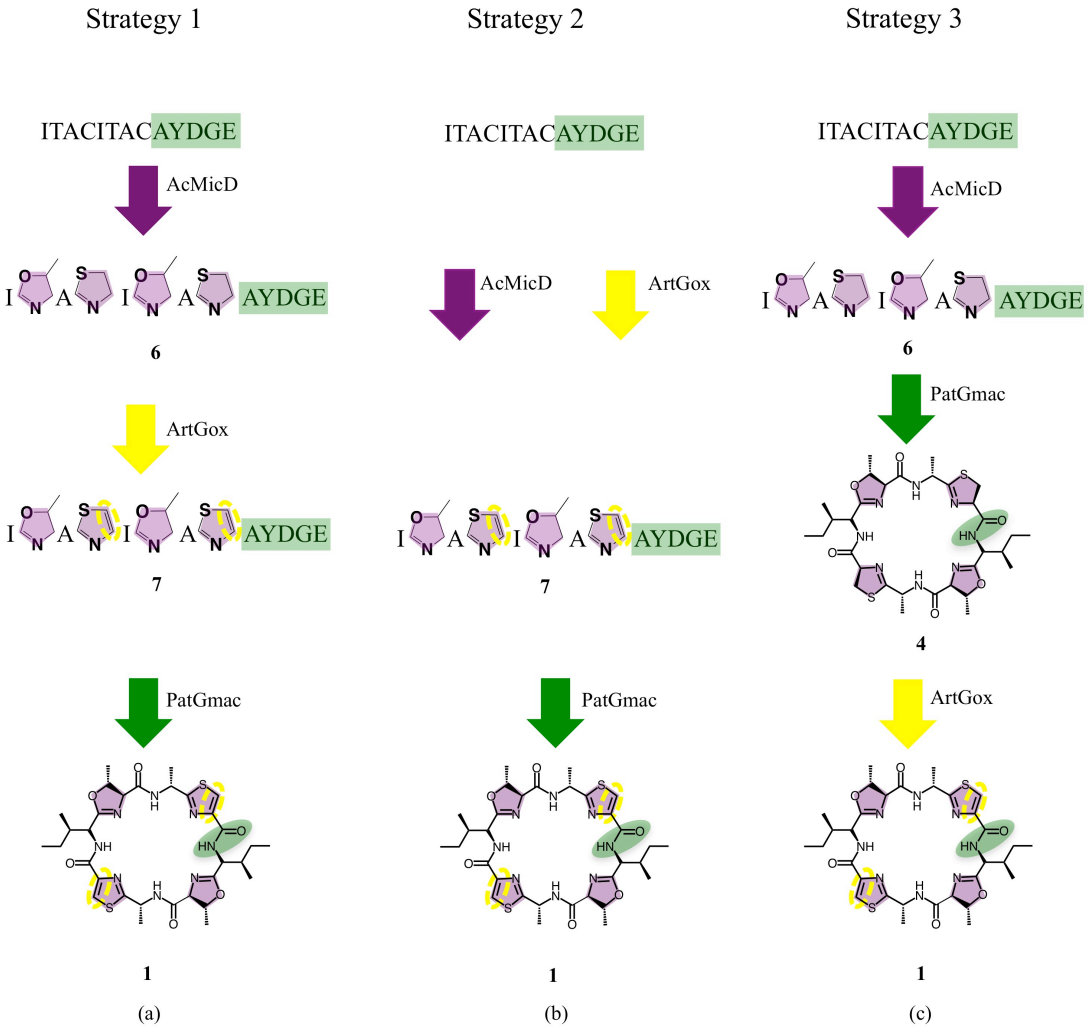


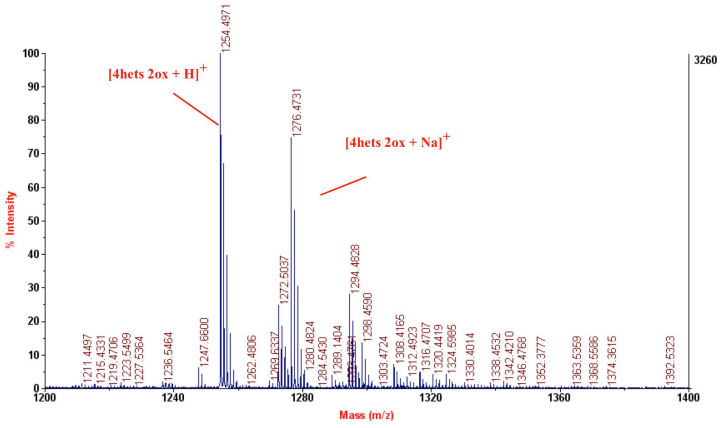

FIGURE 5.9: **Three possible strategies for the one-pot synthesis** (a) Strategy (1) sequential addition of AcMicD, ArtGox and PatGmac. (b) Strategy (2) simultaneous addition of AcMicD, ArtGox, followed by addition of PatGmac. (c) Strategy (3) sequential addition of AcMicD, PatGmac and ArtGox.

For strategies (1) and (2), ArtGox (20 μM) and FMN (200 μM) were added to the reaction mixture (substrate peptide + AcMicD) and incubated at 27 °C for a total reaction time of 32 h (16 h per transformation) in an attempt to generate $IT^{MeOX}AC^{Thz}IT^{MeOx}AC^{Thz}$AYDGE (**7**) from **6** . Reaction progress was monitored using MALDI-TOF-MS (Fig. 5.10).

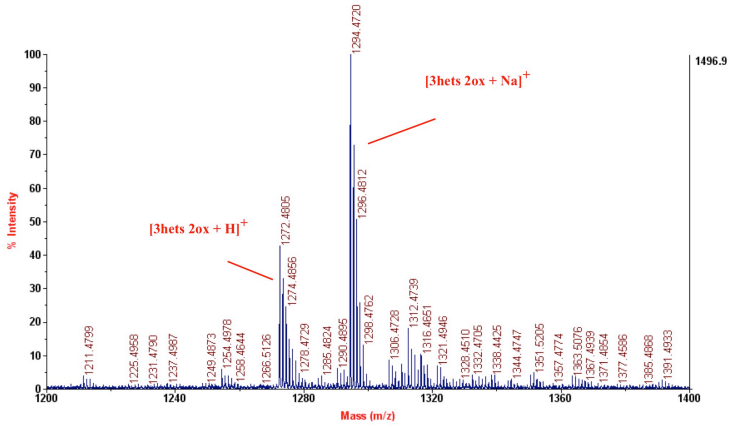
For strategy (1) a mass decrease of 2 Da per thiazole (m/z = 1254.5) relative to the starting material indicated the oxidation reaction was complete (Fig. 5.10). For strategy (2), the completely processed peptide **7** (m/z = 1254.5) was observed, but as a minor product (Fig. 5.10). The major product was identified as a species containing only three heterocycles, two of which were oxidised. *i.e.* both cysteines were processed completely, while one threonine remains unprocessed. (m/z = 1272.5, Fig. 5.10). Prolonged incubation of the reaction did not alter the ratio of the products (data not shown). These data suggest that the rate of heterocyclisation is slow compared with the rate of oxidation and that once a modified peptide has been oxidised it cannot be further processed by the heterocyclase. Therefore, under current reaction conditions, to form an exclusive product, the heterocyclisation reaction must be left to complete before subsequent enzymes are added to the reaction. Therefore strategy (2) was dismissed as a route to the one-pot *in vitro* synthesis of cyanobactins.

For strategies (1) and (3), the macrocyclase enzyme was added to the oxidised and reduced substrates **7** and **6** respectively, as described (section 5.2.4). After 24 h the reaction was monitored using MALDI-TOF-MS (Fig. 5.10). For strategy (3), the reaction had progressed as expected, showing predominantly cyclic product **4** (m/z = 705.4) with some starting material **6** (m/z = 1258.5) present (Fig. 5.10). However for strategy (1), no cyclic peptide product **1** (m/z = 701.3) was observed (Fig. 5.10). After extended incubation (up to 1 week), **1** was observed, but the starting material **7** was still the major species (Fig. 5.10). This suggests that the thiazole containing peptide is a poor substrate of the macrocyclase. Interestingly, when this experiment was repeated using AcLynD in place of AcMicD to generate $ITAC^{Thz}ITAC^{Thz}$AYDGE (**8**), turnover by PatGmac was observed at a rate comparable to that of the two-thiazoline containing peptides (Fig. 5.10). Therefore the ability of PatGmac to process **8**, but not **7**, cannot simply be due to the chemical and electronic differences introduced by the aromatic thiazole, relative to the thiazoline, but due to significant changes in the overall conformation upon thiazole formation, when two additional oxazoline

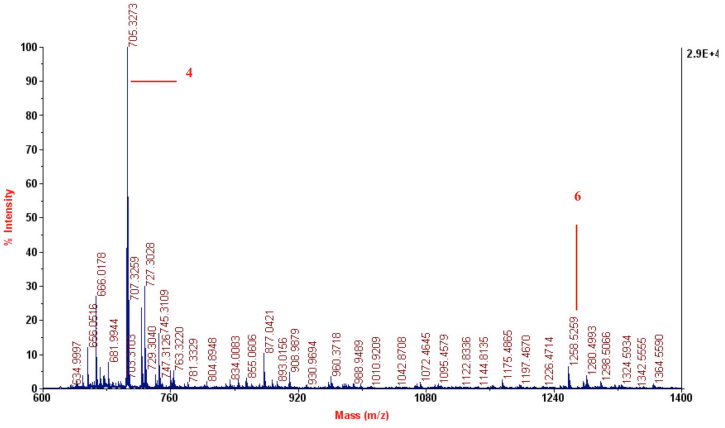
rings are present. Further work towards confirming this hypothesis and understanding cyanobactin oxidation are ongoing, but are not within the scope of this thesis. As strategy (1) is only compatible with peptide substrates processed with AcLynD but not AcMicD, and so is not universally applicable to all substrates, it was also dismissed, and so strategy (3): macrocyclisation before oxidation was deemed to be the order of the enzymatic transformations in a one-pot *in vitro* synthesis.


[4hets 2ox + H]+
[4hets 2ox + Na]+
% Intensity
Mass (m/z)
3260

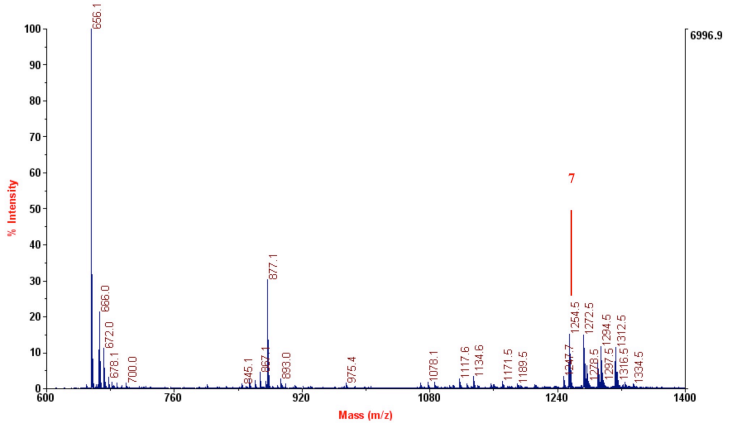
(a)


[3hets 2ox + Na]+
[3hets 2ox + H]+
% Intensity
Mass (m/z)
1496.9

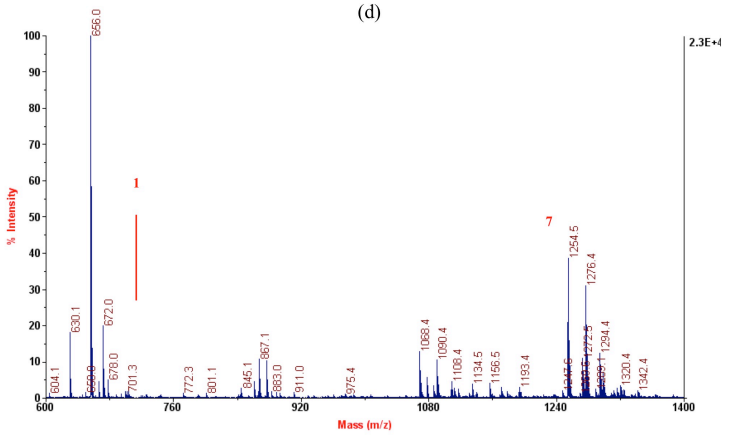
(b)


4
6
% Intensity
Mass (m/z)

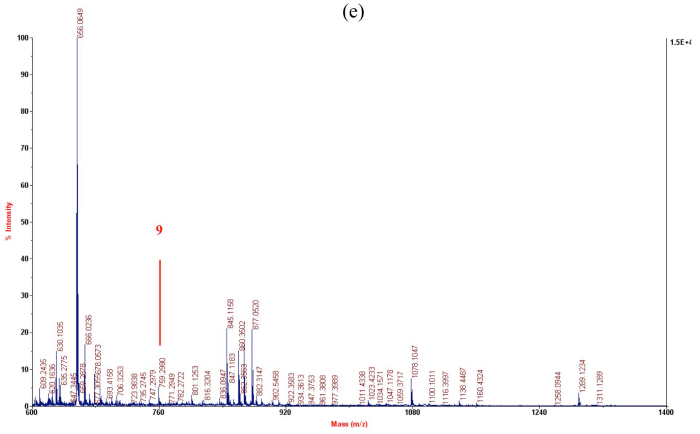
(c)


% Intensity
Mass (m/z)
6996.9
7
666.1
677.1
666.0
672.0
678.1
700.0
845.1
867.1
893.0
975.4
1078.1
1117.6
1134.6
1171.5
1189.5
1247.7
1254.5
1272.5
1294.5
1312.5
1334.5

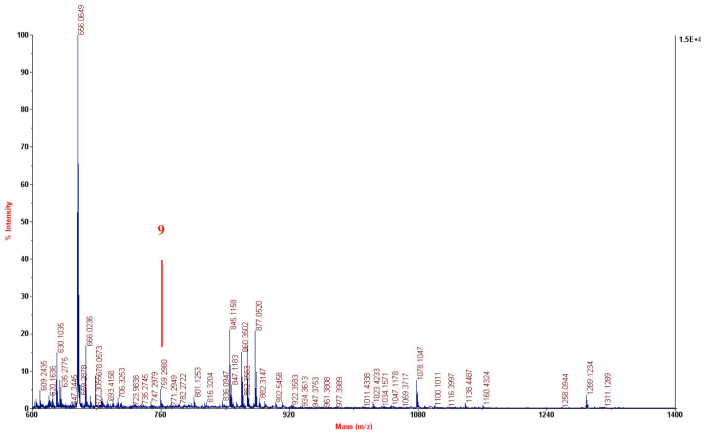
(d)


% Intensity
Mass (m/z)
2.3E+4
1
7
666.0
630.1
672.0
604.1
678.0
701.3
772.3
801.1
845.1
867.1
883.0
911.0
975.4
1068.4
1090.4
1108.4
1134.5
1156.5
1193.4
1254.5
1276.4
1294.4
1320.4
1342.4

(e)


% Intensity
Mass (m/z)
1.5E+4
9
666.0649
630.1035
669.0236
845.1169
880.3932
877.0920
1078.1047
1100.1011
1116.3997
1138.4487
1160.4324
1258.0944
1289.1234
1311.1089

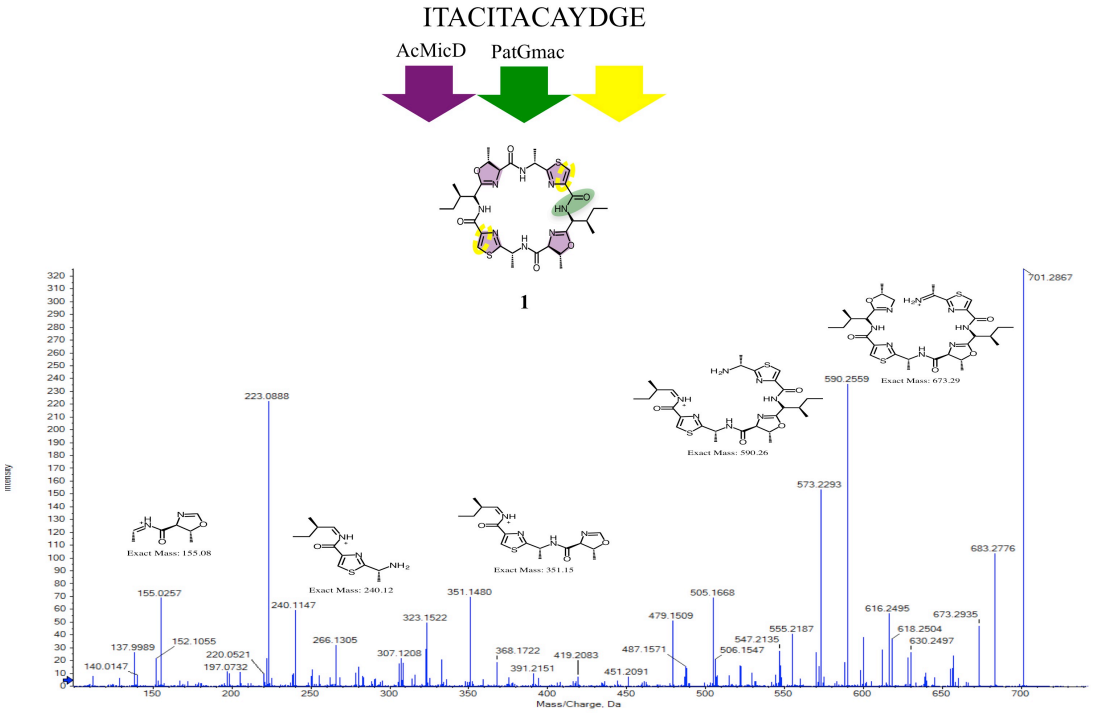
(f)



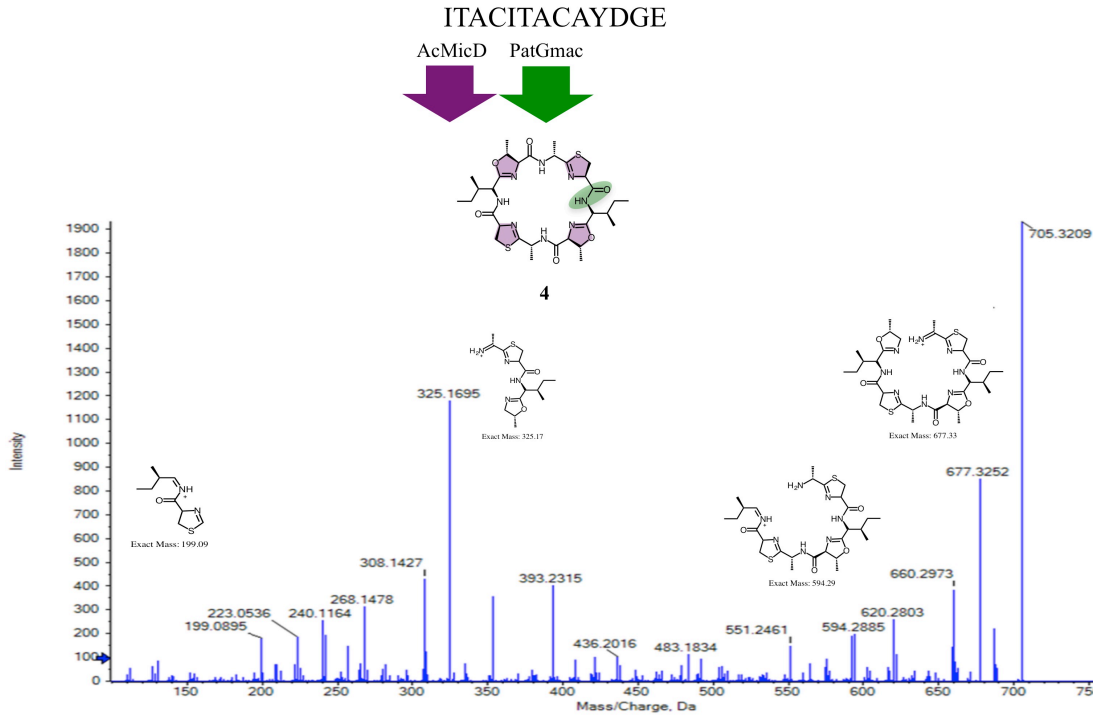

(g)

FIGURE 5.10: **Order of biotransformation reactions** MALDI-TOF-MS of (a) product(s) of heterocyclisation and oxidation reactions as per strategy (1), (b) products of heterocyclisation and oxidation reactions as per strategy (2), (c) macrocyclisation reaction of **6** as per strategy (3) after 24 h. A significant peak of product **4** is observed. (d) macrocyclisation reaction of **7** as per strategy (1) after 24 h. No product **1** is observed. (e) macrocyclisation reaction of **7** as per strategy (1) after 7 d. A hint of product **1** is observed. (f) macrocyclisation of **8** as per strategy (2) after 5 d. Product cyclo[ITAC$^{\mathrm{Thz}}$ITAC$^{\mathrm{Thz}}$] (**9**) is observed; starting material **8** is not. (g) Oxidation of **4** to give **1** as per strategy (1).

As with the original *in vitro* synthesis, addition of ArtGox (20 μM) and FMN (200 μM) to the macrocycle was sufficient to achieve complete oxidation of both thiazolines to thiazoles, forming **1** from **4** following incubation at 27 °C for 16 h, as determined by MALDI-TOF-MS (Fig. 5.10). Confirmation that the final product was the desired **1** was achieved with MSMS analysis (Fig. 5.11). This one-pot *in vitro* synthesis (sequential addition) method was successfully employed without the addition of the oxidase domain, and also using AcLynD, with and without the oxidase, to generate three additional products: Cyclo[IT$^{\mathrm{MeOX}}$AC$^{\mathrm{ThH}}$IT$^{\mathrm{MeOx}}$AC$^{\mathrm{ThH}}$] (**4**) cyclo[ITAC$^{\mathrm{Thz}}$ITAC$^{\mathrm{Thz}}$] (**9**) and cyclo[ITAC$^{\mathrm{ThH}}$ITAC$^{\mathrm{ThH}}$] (**10**) (Fig. 5.11).


ITACITACAYDGE
AcMicD
PatGmac
1
Intensity
Mass/Charge, Da
701.2867
590.2559
573.2293
683.2776
223.0888
351.1480
505.1668
323.1522
240.1147
155.0257
479.1509
616.2495
673.2935
555.2187
618.2504
630.2497
547.2135
506.1547
487.1571
266.1305
307.1208
368.1722
419.2083
391.2151
451.2091
137.9989
152.1055
220.0521
197.0732
140.0147
Exact Mass: 673.29
Exact Mass: 590.26
Exact Mass: 155.08
Exact Mass: 240.12
Exact Mass: 351.15

(a)


ITACITACAYDGE
AcMicD
PatGmac
4
Intensity
Mass/Charge, Da
705.3209
677.3252
325.1695
308.1427
393.2315
660.2973
268.1478
223.0536
240.1164
199.0895
620.2803
594.2885
551.2461
436.2016
483.1834
Exact Mass: 677.33
Exact Mass: 325.17
Exact Mass: 199.09
Exact Mass: 594.29

(b)

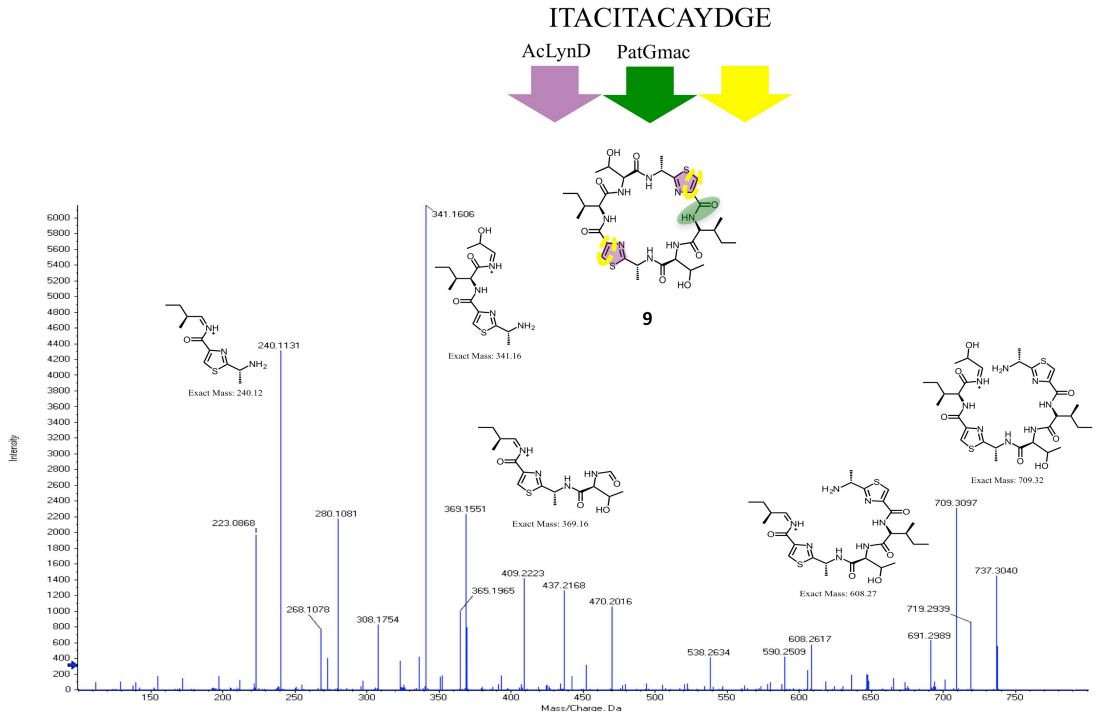
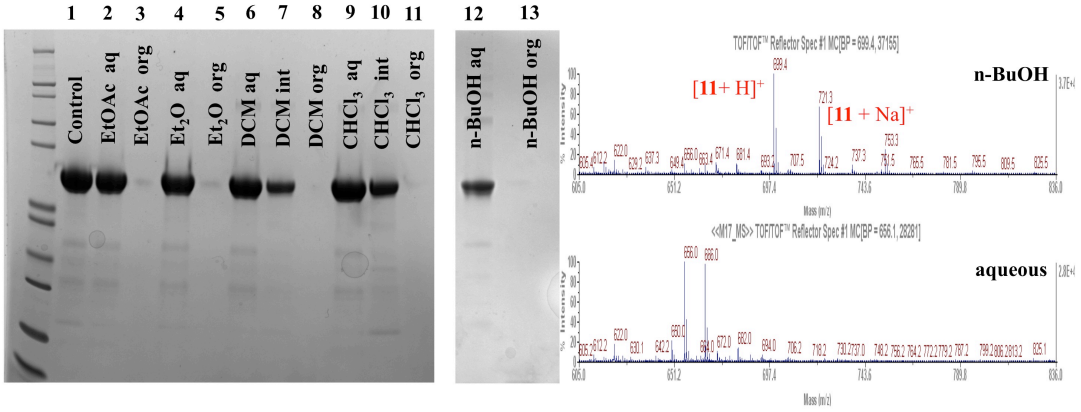



FIGURE 5.11: **One-pot *in vitro* synthesis products** MSMS fragmentation of (a) **1** (b) **4** (c) **9** (d) **10**.

### 5.3.3 Purification of One-pot *in vitro* Synthesis Products

The one-pot *in vitro* synthesis products require purification from a complex mixture of proteins, small molecules, salts and other peptides. The high concentration of multiple enzymes in the reaction mixture, coupled with a low concentration of the product ($\leq$ 90 μM) make direct HPLC purification of the reaction mixture impractical. Accordingly, a rapid, robust and straightforward isolation of the cyclic peptide product(s) from the enzyme mixture, and subsequent concentration of the product, is required before HPLC purification. The chemical composition of cyanobactins make them characteristically hydrophobic, and so we reasoned a simple two-phase extraction method might offer such a solution, if an organic solvent, which when mixed with the reaction mixture selectively extracts and retains the cyclic peptides from the aqueous reaction mixture, can be found. In search of a suitable extraction solvent, a test reaction was set up, which for simplicity concerned only the macrocyclisation reaction. The substrate peptide VGAGIGFPAYDG (made in-house by Dr Brunello Nardone) was incubated with PatGmac. The completed reaction was divided into five samples, and each sample was extracted with a different solvent: ethyl acetate (EtOAc), diethyl ether ($Et_2O$), dichloromethane (DCM), chloroform ($CHCl_3$) and n-butanol (n-BuOH). SDS-PAGE analysis confirmed that all solvents successfully retained PatGmac in the aqueous phase (Fig. 5.12). However only n-BuOH could extract all the cyclo[VGAGIGFP] (**11**) into the organic phase, as determine by MALDI-TOF-MS (Fig. 5.12).

To examine the viability of n-BuOH as an extraction solvent for the initial isolation of one-pot *in vitro* synthesis products from more complicated reaction mixtures, the extraction experiment was repeated with product **1**. The reaction mixture and n-BuOH were mixed in a 1:1 ratio. Upon mixing a large emulsion formed at the interface between the two phases. Centrifuging the sample at 4000$g$ for 10 min reduced the emulsion and separated the two phases. The two phases were pooled separately and the aqueous phase was extracted a further two times with equal volumes of n-BuOH, the organic phase was pooled and kept separately in each case. Analysis by MALDI-TOF-MS confirmed **1** to be in the organic phase, and was completely absent from the aqueous phase (Fig. 5.13).




(a) (b)

FIGURE 5.12: **Extraction of cyclic peptide products** (a) SDS-PAGE of PatGmac after extraction of **11** using various solvents - the aqueous (aq) and organic (org) phases were analysed for each solvent. For DCM and $CHCl_3$ samples of the emulsion at the interface between the two phases (Int) were also analysed. (b) MALDI-TOF-MS of organic (top) and aqueous (bottom) phases following extraction of **12** (mW = 698.4 Da) from the reaction mixture using n-BuOH.

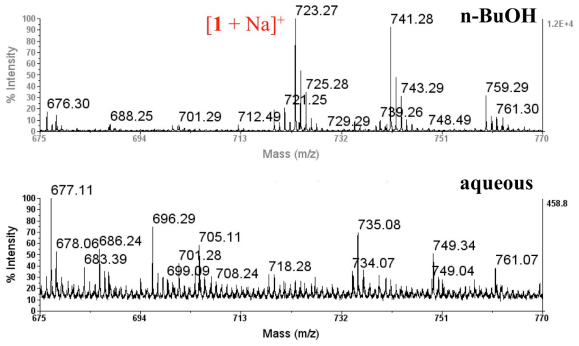
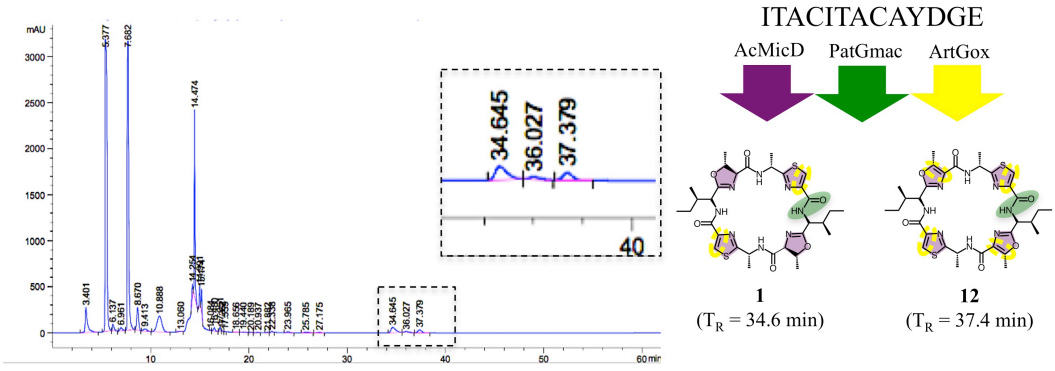



FIGURE 5.13: **Extraction of 1** MALDI-TOF-MS of organic (top) and aqueous (bottom) phases following extraction of **1** (mW = 700.3 Da) from the reaction mixture n-BuOH. The peak at m/z = 723.3 in the organic phase, and accordingly its absence from the aqueous phase indicates **1** is successfully extracted from the reaction mixture using n-BuOH. The peak at m/z = 741.3 in the n-BuOH fraction corresponds to a species containing three heterocycles, likely arising due to acid catalysed heterocycle ring opening under MALDI conditions.

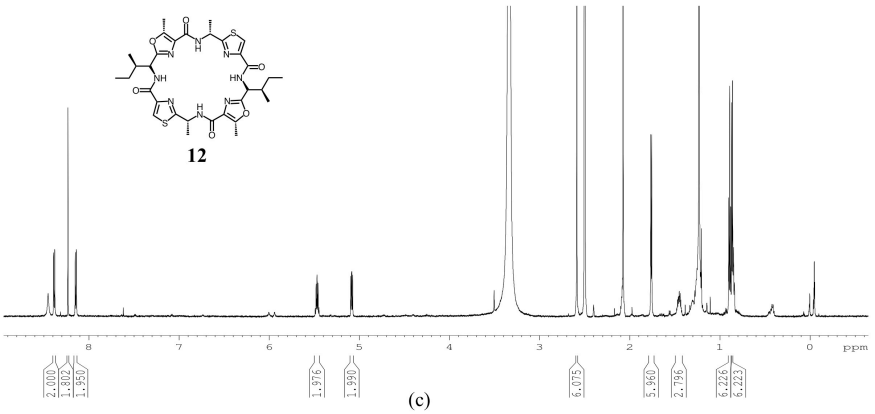
For all products made to date in this lab, products were always found in the organic phase from the first extraction, and sometimes in the organic phase from the second third extractions. After three extractions, macrocyclic products were never identified in the aqueous phase.

The organic phase containing **1** was dried by rotary evaporation and re-solubilised in

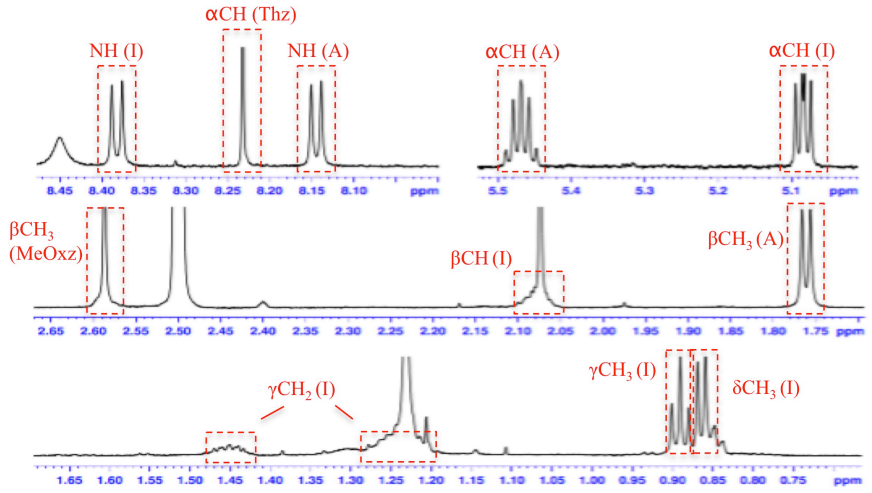
an equal volume of methanol and water for HPLC purification as described (section 5.2.6). Compound **1** eluted from the RP-HPLC C18 column as a single peak at retention time ($T_R$) of 34.6 min, as identified by MS (Fig. 5.14). A second peak, at $T_R$ = 37.4 min was observed, and corresponds to an unknown compound of m/z = 697.2. It was suspected this compound was cyclo[$IT^{MeOXz}AC^{Thz}IT^{MeOxz}AC^{Thz}$] (**12**), a quadruple oxidised product of the ArtGox reaction, where the oxazolines were oxidised to oxazoles. Attempts to acquire MSMS data were intractable, so the identity of **12** was instead confirmed by NMR (Fig. 5.14). Due to a limited amount of material, a $^{13}C$ NMR spectrum was not obtained and so the quaternary carbons have not been assigned. Despite this we are confident the identity of **12** is correct. The presence of **12** was surprising for two reasons: firstly, no hint of **12** had previously been identified by MALDI-TOF-MS (Fig. 5.10); secondly, while natural cyanobactins containing oxazoles as well as thiazoles exist[64,65], no products of the arthrospiramide pathway (where from ArtGox originates) contain oxazolines or oxazoles[112], and so ArtGox was not expected to catalyse oxazoline oxidation. Instead, it was suspected the activity of the cyanobactin oxidase enzymes reflected that observed for the heterocyclases, where different enzymes are responsible for different chemistries, and so it was thought a different enzyme would be needed to achieve this transformation (section 1.4.3). The ratio of the products **1** and **12** of approximately 70:30 (Fig. 5.14) indicate that oxazoline oxidation is not an occasional side reaction, and that by adjusting the conditions of the oxidation reaction it should be possible to drive the reaction towards oxazole formation, making **12** as an exclusive product. Similarly, the reaction conditions must also be altered to favour only thiazoline oxidation, in order to generate **1** as an exclusive product, thus maximising the yield of the desired product in each case. Importantly, if this oxidation reaction can be controlled to give macrocycles containing thiazoles, or thiazoles and oxazoles, it increases the chemical diversity accessible through the one-pot *in vitro* synthesis, without the need for additional enzymes.


mAU
3000
2500
2000
1500
1000
500
0
3.401
6.377
6.961
7.682
8.670
9.413
10.888
13.050
14.474
23.965
25.785
27.175
34.645
36.027
37.379
10
20
30
40
50
60 min
34.645
36.027
37.379
40
ITACITACAYDGE
AcMicD
PatGmac
ArtGox
1
(T$_R$ = 34.6 min)
12
(T$_R$ = 37.4 min)

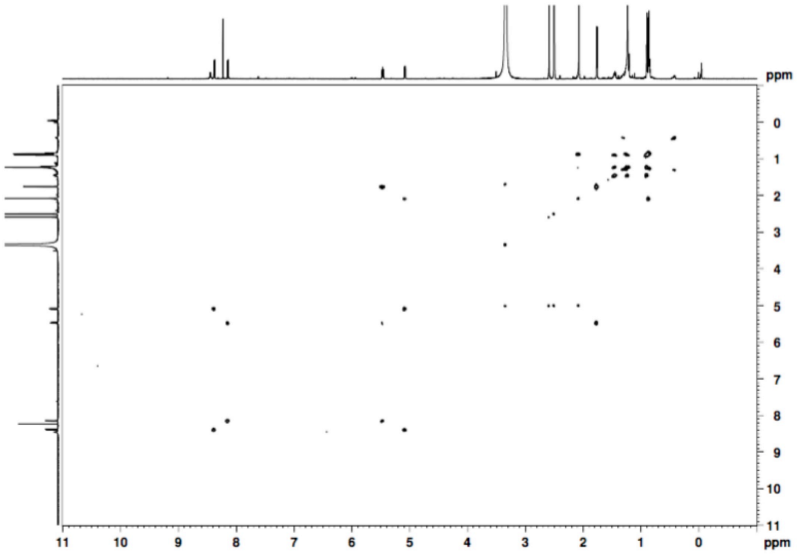
(a) (b)


12
8
7
6
5
4
3
2
1
0
ppm
2.000
1.802
1.950
1.976
1.990
6.075
5.960
2.796
6.226
6.223

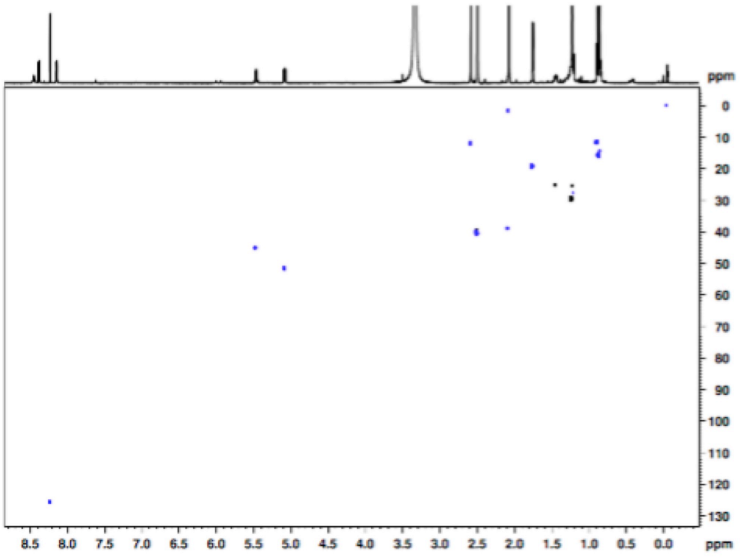
(c)


αCH (Thz)
NH (I)
NH (A)
αCH (A)
αCH (I)
8.45 8.40 8.35 8.30 8.25 8.20 8.15 8.10 ppm
5.5 5.4 5.3 5.2 5.1 ppm
βCH$_3$ (MeOxz)
βCH (I)
βCH$_3$ (A)
2.65 2.60 2.55 2.50 2.45 2.40 2.35 2.30 2.25 2.20 2.15 2.10 2.05 2.00 1.95 1.90 1.85 1.80 1.75 ppm
γCH$_2$ (I)
γCH$_3$ (I)
δCH$_3$ (I)
1.65 1.60 1.55 1.50 1.45 1.40 1.35 1.30 1.25 1.20 1.15 1.10 1.05 1.00 0.95 0.90 0.85 0.80 0.75 ppm

(d)




(e)




(f)

FIGURE 5.14: **Characterisation of one-pot products** RP-HPLC purification of **1** extracted from one-pot synthesis reaction. Product elutes at $T_R = 34.6$ min as identified by MS. A second one-pot synthesis ($T_R = 37.4$ min; m/z = 697.2) thought to be an alternative reaction product **12**. Insert shows a zoom of the product region. (b) One-pot reaction scheme showing possible products formed under these reaction conditions. (c) $^1$H NMR spectrum of **12**. Peak integrals are shown below the spectrum. (d) zoom of $^1$H NMR spectrum of **12** showing assigned peaks. (e) COSY NMR spectrum of **12**. (f) $^1$H-$^{13}$C HSQC spectrum of **12**.

TABLE 5.1: **$^{1}$H NMR, COSY, $^{1}$H-$^{13}$C HSQC analysis (500 MHz, MeOH) of 12.**

| Amino acid | Atom | $^{1}$H chemical shift | $^{13}$C chemical shift |
|---|---|---|---|
| Ile | NH | 8.39, d, J = 8.7 | - |
| | αCH | 5.08, m | 51.5 |
| | βCH | 2.09, m | 38.9 |
| | $\beta CH_2$ | 1.46, m; 1.25 m | 24.9 |
| | $\gamma CH_3$ | 0.86, d, J = 6.9 | 15.3 |
| | $\delta CH_3$ | 0.89, t, J = 7.4 | 11.4 |
| MeOxz | $\beta CH_3$ | 2.59, s | 11.6 |
| Ala | NH | 8.14, d, J = 8.3 | - |
| | αCH | 5.47, m | 44.9 |
| | $\beta CH_3$ | 1.76, d, J = 6.9 | 19.0 |
| Thz | CH | 8.23, s | 125.5 |

### 5.3.4 Controlling the Oxidation Reaction

To try to determine ArtGox reaction conditions that result in only thiazoline oxidation, the rate of the oxidation reaction was examined in more detail. Macrocycle **4** was incubated with 5, 10 or 20 μM ArtGox, and 200 μM FMN, and the reaction was monitored every hour by MALDI-TOF-MS. Both thiazolines were completely oxidised (-4 Da) after only 3 h with 10 and 20 μM ArtGox, and after 4 h with 5 μM ArtGox, indicating the initial reaction conditions of 20 μM ArtGox for 16 h were excessive and unnecessary (data not shown). Analysis of **4** incubated with 5 μM ArtGox and 200 μM FMN for 4 h via HPLC-MS revealed these reaction conditions resulted in only the thiazole containing product **1**, with a peak observed at $T_R$ = 34.6 min, and no hint of the oxazole containing product detected (Fig. 5.15).

To encourage oxazole formation, **4** was incubated with 40 μM ArtGox and 1 mM FMN for 48 h. This time, using HPLC-MS, the major product was identified as the oxazole-containing macrocycle **12** (Fig. 5.15). To determine whether it was the extra enzyme, the extra FMN, or the longer incubation time that was the key driving force for oxazole formation, four new oxidation reactions of **4** were set up under the following conditions:

(1) 20 μM ArtGox, 200 μM FMN; (2) 20 μM ArtGox, 1 mM FMN; (3) 40 μM ArtGox, 200 μM FMN; (4) 40 μM ArtGox, 1 mM FMN. The reactions were monitored by HPLC after 24 and 48 h. After 24 h reaction (1) contained approximately a 70:30 mixture of **1** and **12**, while the ratio observed for reactions (2), (3) and (4) were 65:35, 25:75 and 25:75 respectively (Fig. 5.15). After 48 h the ratio of products **1** to **12** in reactions (1-4) were 60:40, 20:80, 5:95 and 5:95 respectively. These data show that simply extending the incubation time (48 h vs 24 h) is not sufficient to drive oxazoline oxidation under standard conditions (20 μM ArtGox, 200 μM FMN). Extra FMN (1 mM vs 200 μM) does contribute to oxazoline oxidaiton, but not as significantly as extra enzyme (40 μM vs 20 μM). The requirement of extra enzyme to achieve complete oxazoline oxidation would suggest that oxazoline is a poor substrate for the enzyme. In summary, it would be prudent to adopt conditions of reaction (4) as the strategy to drive oxidation of oxazoline to completion, and generate **12** as an exclusive product. In this way, by varying the oxidation reaction conditions, in conjunction with the use of either AcMicD or AcLynD, it is possible, using the one-pot *in vitro* synthesis method, to generate five different products from a single starting peptide enabling easy, cost-effective diversification.

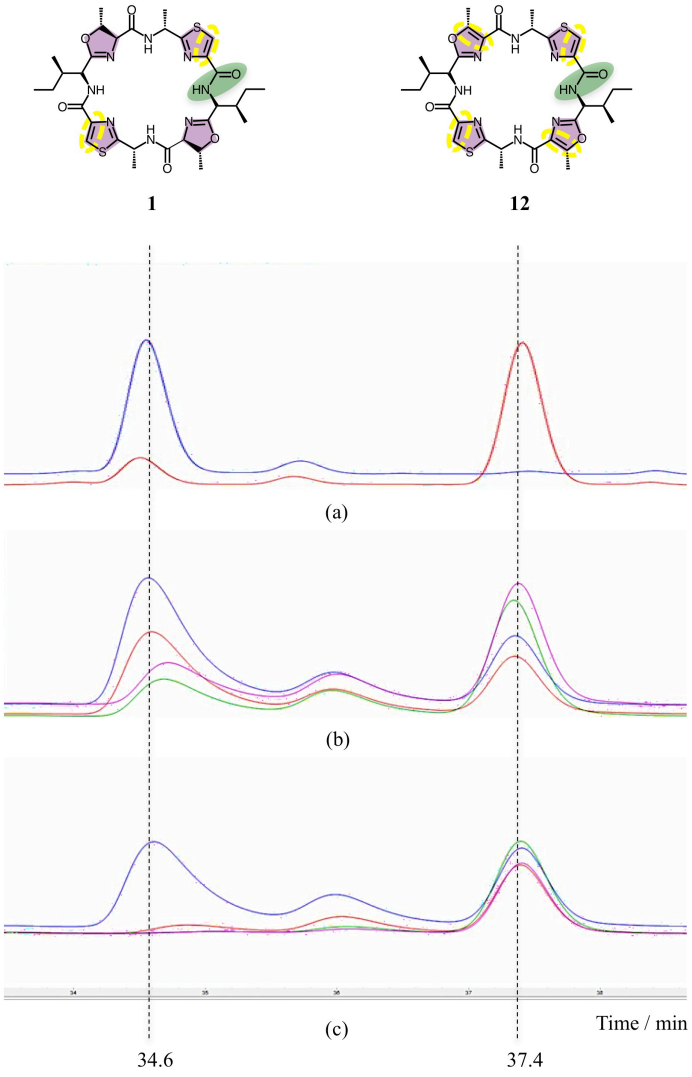



FIGURE 5.15: **Controlling the oxidation reaction** (a) RP-HPLC chromatograms of **4** subjected to various oxidation reaction conditions to form either **1** or **12**. (a) 5 µM ArtGox, 200 µM FMN, 4 h (blue); 40 µM ArtGox, 1 mM FMN, 48 h (red). (b) reaction 1, 24 h (blue); reaction 2, 24 h (red); reaction 3, 24 h (green); reaction 4, 24h (pink). (c) reaction 1, 48 h (blue); reaction 2, 48 h (red); reaction 3, 48 h (green); reaction 4, 48 h (pink).

### 5.3.5 Quantitation of One-pot *in vitro* Synthesis Products

Quantitation of the cyclic peptide products of the one-pot *in vitro* synthesis is a non-trivial task. The retention of water and buffer salts when peptides are vacuum or freeze-dried to a powder mean that the quantity and yield of the pure product cannot simply be determined using an accurate mass balance[113]. Furthermore natural cyanobactins, and thus many of the desired analogues do not contain tryptophan residues, precluding quantitation using the Beer-Lambert law at $A_{280}$. An alternative method would be

to quantify products by HPLC compared to standard curve, but this is unfeasible due to the lack of authentic standard. One solution is to quantify the pure cyclic peptide products by NMR. Quantitative NMR (qNMR) relies on the fact that the intensity of a signal (a peak in an NMR spectrum) is directly proportional to the number of nuclei evoking the signal, and thus the concentration of the sample[114]. The concentration of the analyte in question can be determined through comparison with an internal reference standard of known concentration[115]. Importantly the standard to be used as the internal or external reference is not an authentic standard, which might be expensive, of limited availability or does not exist; but small molecules readily available in high purity, easily weighable, highly soluble in different NMR solvents, and provide a unique, distinct set of signals[115]. Development of the ERETIC method (Electronic REference To access In vivo Concentrations) provides a reference signal electronically, and so the reference standard can be measured externally of the analyte sample, allowing the sample to be recovered[116]. Thus qNMR provides a non-destructive, rapid and universally applicable method of quantitation, which can form part of standard structural characterisation protocols. qNMR is used routinely in the pharmaceutical industry[114] and has previously been used to quantify cyanobactin production *in vivo*[57].

To validate the use of qNMR for the one-pot *in vitro* synthesis, cyclo[VGAGIGWP] (**13**) was synthesised and purified and described, from the starting peptide VGAGIG-WPAYDG (made in-house by Dr Brunello Nardone). The inclusion of a tryptophan residue allows for comparison between the concentration of **13** as calculated by qNMR and that determined by UV-absorbance at 280 nm.

HPLC-purified **13** (95 %) was freeze-dried to give 5.6 mg of white powder. The entire sample was solubilised in 500 µL d6-DMSO. Using a theoretical extinction coefficient of 5 500 $M^{-1}$ $cm^{-1}$ (as calculated by *ExPASy ProtParam* part of the *ExPASy server*[117]) the concentration of **13** was determined to be 13.74 mM, equating to 5.063 mg of compound. The solution of **13** was then diluted to 600 µL with d6-DMSO for qNMR analysis. A solution of 10 mM anthrinalic acid (**14**) was prepared as an external reference standard. qNMR spectra were recorded for both **13** and **14**. An ERETIC reference of 10 mM was generated by integrating one of the well-defined and isolated benzyl-hydrogen peaks (correlating to 1H) from the qNMR spectrum of **14** (Fig. 5.16). The concentration of **13** was then calculated by integrating the distinct peak of the indole nitrogen (1H) from its qNMR spectrum and comparing the value with that of the ERETIC reference. This

gave a concentration of **13** equal to 11.49 mM, (Fig. 5.16) which corresponds to 5.081 mg. The total amount of **13**, as determined by UV-absorbance and qNMR are in good agreement with each other, giving 5.063 and 5.081 mg respectively, suggesting qNMR can be relied upon as a method to quantify cyclic peptides and cyanobactin analogues made via one-pot *in vitro* synthesis.

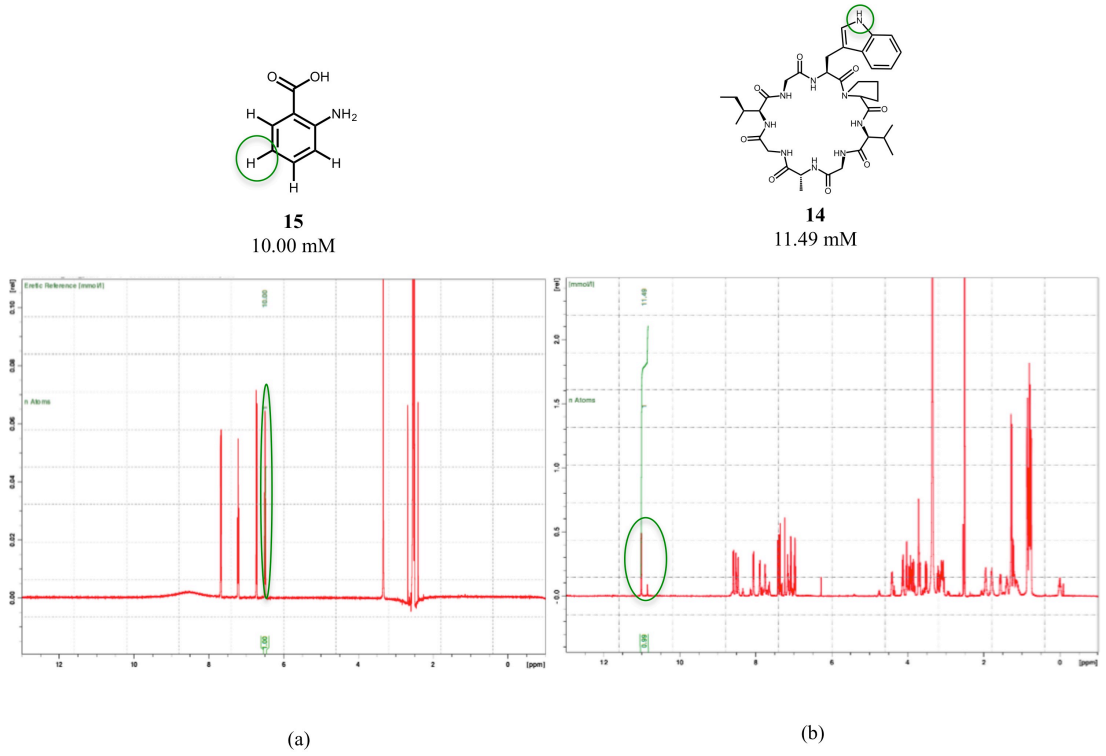



FIGURE 5.16: **Quantitation of 13** qNMR spectra of (a) **14** and (b) **13**. The proton used for quantitation is shown on the structure and in the qNMR spectra in green. N.B. **13** was found to exist as two conformers, giving rise to doubling of peaks. Due to this, both indole peaks (as highlighted) were integrated together for accurate quantitation.

### 5.3.6 One-pot *in vitro* Synthesis of Patellamide D Analogues to Determine Structure-activity Relationships of Pgp Inhibition

In collaboration with Dr Andrew Bent, Scottish Enterprises and Ripptide-Pharma, a series of patellamide D analogues were designed in an attempt to determine structure-activity relationships (SAR) governing Pgp inhibition. A diversification strategy was designed to investigate two key structural elements, which are conserved in natural patellamdes, by varying (1) the number and type of heterocycles; and (2) the intervening amino acids between the heterocycles at positions 1, 3, 5 and 7 (Fig. 5.17).

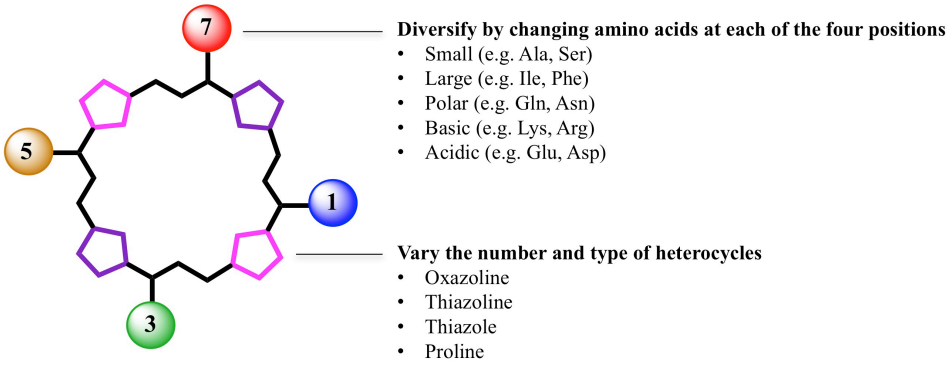



FIGURE 5.17: **SAR diversification strategy** By varying the number and type of heterocycles, and the interveneing amino acids, the SAR for patellamide mediated Pgp inhibition can be investigated.

Therefore positions 2 and 6 were fixed to be either serine/threonine or proline, while positions 4 and 8 were fixed to be either a cysteine or a proline residue. Any one or two of positions 1, 3, 5 and 7 were then varied. Using such a strategy, ten starting peptides were designed and processed with different enzyme combinations to give twenty-one discrete macrocyclic products, including patellamide D as a positive control and one linear product to act as a negative control. All starting peptides were ordered from BioSyn. The starting peptides, enzyme transformations and the resulting final products can be seen in Table 4.2. Structures of all the compounds can be seen in Fig. 5.18. All the compounds listed in Table 4.2 and Fig. 5.18 were solubilised in the minimum possible volume of d6-DMSO, and identity was confirmed by MS and MSMS, purity assessed by HPLC and quantities determined by qNMR. Data can be seen in Appendix B.

N. B. Accurate quantitation by NMR requires the NMR spectra to be assigned, and thus the NMR must be clearly resolved and of good quality. The qNMR for the compounds, as seen in Appendix B, are of variable quality and so the quantitation can only be considered approximate. Although only approximate, this quantitation method is preferred to quantitation based on the amount of dried solid, which contain varying amounts of water, which in our experience in the lab range from 10 - 60 %. Peaks deemed suitable for quantitation were identified tentatively based on the chemical shift, splitting pattern, and integration of the $^{1}$H NMR spectra, and those used for each compound can be seen in Appendix B.

TABLE 5.2: **Compounds for SAR**

| **Starting peptide sequence** | **Biotransformation enzymes** | **Product** |
|---|---|---|
| ITACITFCAYD | AcLynD, PatGmac | cyclo[ITAC$^{\mathrm{ThH}}$ITFC$^{\mathrm{ThH}}$] (**15**) |
| | AcLynD, PatGmac, ArtGox | cyclo[ITAC$^{\mathrm{Thz}}$ITFC$^{\mathrm{Thz}}$] (**16**) |
| | AcPatD, PatGmac, | cyclo[IT$^{\mathrm{MeOx}}$AC$^{\mathrm{ThH}}$IT$^{\mathrm{MeOx}}$FC$^{\mathrm{ThH}}$] (**17**) |
| | AcPatD, PatGmac, ArtGox | cyclo[IT$^{\mathrm{MeOx}}$AC$^{\mathrm{Thz}}$IT$^{\mathrm{MeOx}}$FC$^{\mathrm{Thz}}$] (patellamide D; **18**) |
| ITACIPFCAYD | AcLynD, PatGmac | cyclo[ITAC$^{\mathrm{ThH}}$IPFC$^{\mathrm{ThH}}$] (**19**) |
| | AcLynD, PatGmac, ArtGox | cyclo[ITAC$^{\mathrm{Thz}}$IPFC$^{\mathrm{Thz}}$] (**20**) |
| ITNCITACAYD | AcMicD, PatGmac | cyclo[ITNC$^{\mathrm{ThH}}$IT$^{\mathrm{MeOx}}$AC$^{\mathrm{ThH}}$] (**21**) |
| | AcMicD, PatGmac, ArtGox | cyclo[ITNC$^{\mathrm{Thz}}$IT$^{\mathrm{MeOx}}$AC$^{\mathrm{Thz}}$] (**22**) |
| ITACITFPAYD | AcLynD, PatGmac | cyclo[ITAC$^{\mathrm{ThH}}$ITFP] (**23**) |
| IPACIPFCAYD | AcLynD, PatGmac | cyclo[IPAC$^{\mathrm{ThH}}$IPFC$^{\mathrm{ThH}}$] (**24**) |

| | | |
|---|---|---|
| | AcLynD, PatGmac, ArtGox | cyclo[IPAC$^{\mathrm{Thz}}$IPFC$^{\mathrm{Thz}}$] (**25**) |
| ITACITACAYD | AcLynD, PatGmac | cyclo[ITAC$^{\mathrm{ThH}}$ITAC$^{\mathrm{ThH}}$] (**10**) |
| | AcLynD, PatGmac, ArtGox | cyclo[ITAC$^{\mathrm{Thz}}$ITAC$^{\mathrm{Thz}}$] (**9**) |
| | AcMicD, PatGmac | cyclo[IT$^{\mathrm{MeOx}}$AC$^{\mathrm{ThH}}$IT$^{\mathrm{MeOx}}$AC$^{\mathrm{ThH}}$] (**4**) |
| | AcMicD, PatGmac, ArtGox | cyclo[IT$^{\mathrm{MeOx}}$AC$^{\mathrm{Thz}}$IT$^{\mathrm{MeOx}}$AC$^{\mathrm{Thz}}$] (**1**) |
| ISACITACAYD | AcLynD, PatGmac | cyclo[ISAC$^{\mathrm{ThH}}$ITAC$^{\mathrm{ThH}}$] (**26**) |
| | AcLynD, PatGmac, ArtGox | cyclo[ISAC$^{\mathrm{Thz}}$ITAC$^{\mathrm{Thz}}$] (**27**) |
| IPACITFCAYD | AcMicD, PatGmac | cyclo[IPAC$^{\mathrm{ThH}}$IT$^{\mathrm{MeOx}}$FC$^{\mathrm{ThH}}$] (**28**) |
| | AcMicD, PatGmac, ArtGox | cyclo[IPAC$^{\mathrm{Thz}}$IT$^{\mathrm{MeOx}}$FC$^{\mathrm{Thz}}$] (**29**) |
| ITACHTFCAYD | AcLynD, PatGmac | cyclo[ITAC$^{\mathrm{ThH}}$HTFC$^{\mathrm{ThH}}$] (**30**) |
| | AcLynD, PatGmac, ArtGox | cyclo[ITAC$^{\mathrm{Thz}}$HTFC$^{\mathrm{Thz}}$] (**31**) |
| IPAPIPFPAYD | PatGmac | IPAPIPFP (**32**) |

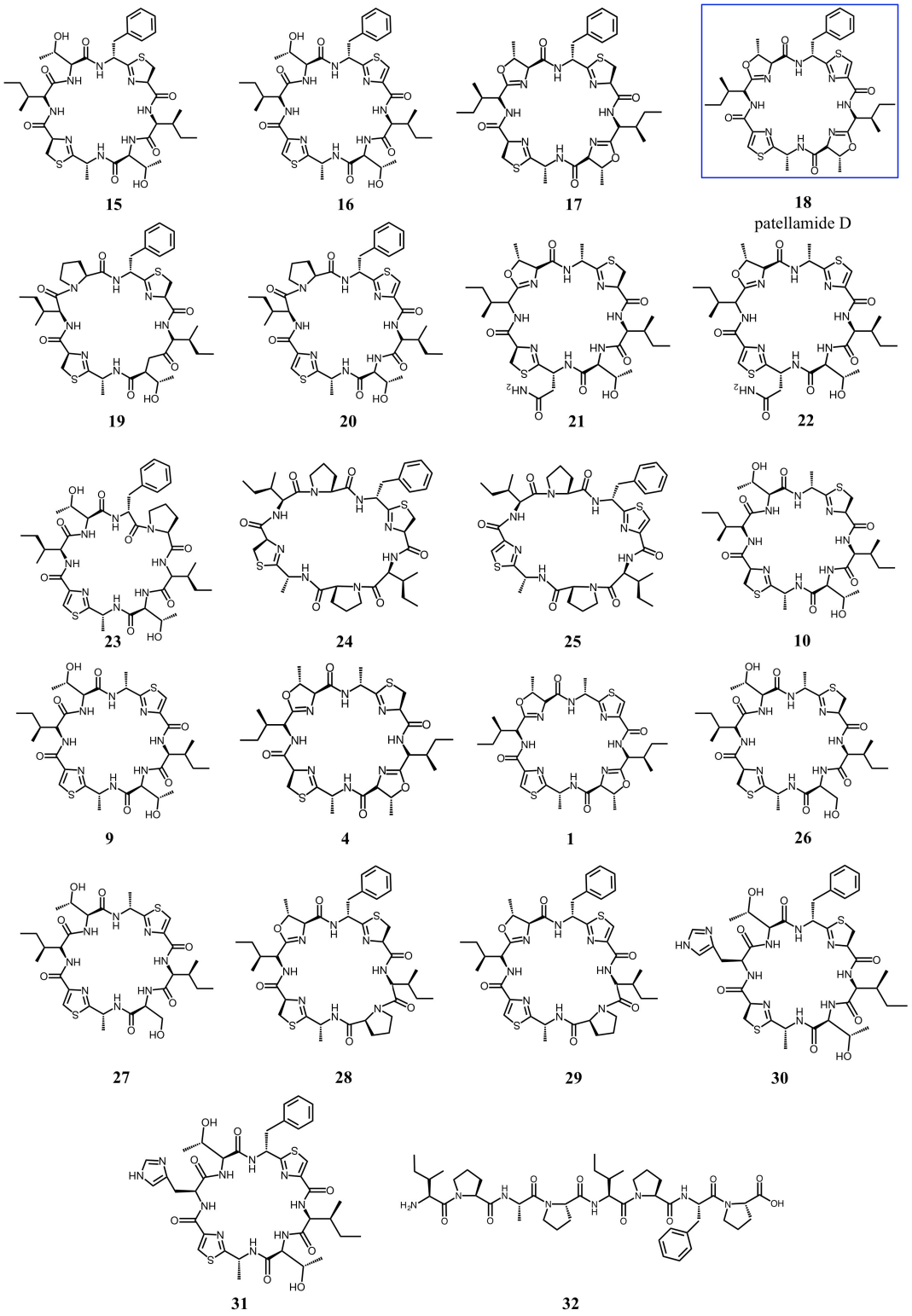



FIGURE 5.18: **SAR compound library** Molecular structures of compounds **1**, **4**, **9**, **10** and **15-32** made via one-pot *in vitro* synthesis to investigate SAR of Pgp inhibition. The natural compound, patellamide D (**18**) is highlighted by a blue box.

Once characterised, the twenty-two compounds listed in Fig. 5.18 and Table 4.2 were subjected to the Pgp-Glo$^{TM}$ assay (Promega). The assay was first validated using a known Pgp inhibitor, cyclosporin A (CsA) as a positive control[31]. For this initial experiment, five Pgp + ATP reactions were prepared: (1) Pgp only, (2) Pgp + 100 μM $Na_3VO_4$, (3) Pgp + 200 μM verapamil, (4) Pgp + 50 μM CsA and (5) Pgp + 200 μM verapamil + 50 μM CsA. Each reaction was set up in triplicate and an average luminescence signal for each reaction was calculated. The relative ATPase activity was determined by subtracting the average luminescence signal for each reaction from that calculated for the $Na_3VO_4$ control sample; ergo the ATPase activity of the $Na_3VO_4$ reaction is 0 (Fig. 5.19). In this way a basal ATPase activity of the untreated Pgp membrane fraction of 180 relative light units (RLU) was determined. Addition of 200 μM verapamil increased the Pgp ATPase activity approximately 4 fold. Addition of 50 μM CsA in isolation did not affect the ATPase activity, however when both 50 μM CsA and 200 μM verapamil were added to the Pgp reaction, the verapamil stimulated ATPase activity was significantly reduced, returning to approximately basal level, confirming CsA inhibits the transport of verapamil across the membrane by Pgp. By varying the concentration of CsA in the Pgp + 200 μM verapamil reaction, a dose response curve was generated and an $IC_{50}$ = 4.09 +/- 1.25 μM was calculated (Fig. 5.19), which is consistant with literature reported values[118–120].

To further validate the assay, the experiment was repeated using **18** (patellamide D) in place of CsA (Fig. 5.19). As before the addition of 200 μM verapamil to Pgp resulted in a 4 fold increase in ATPase activity compared with the untreated sample. This time, the addition of 50 μM **18** in isolation was mildly stimulatory, and increased the ATPase activity approximately 1.5 fold. However the ATPase activity in response to **18** was unchanged in the presence of verapamil, indicating **18** also significantly inhibits verapamil-stimulated ATPase activity. The dose-dependant inhibition of Pgp by **18** was measured, giving an $IC_{50}$ = 8.53 +/- 1.22 μM (Fig. 5.19), which is in good agreement with previous observations[24,25].

To ensure that the observed inhibition of Pgp by both CsA and patellamide D, two structurally unrelated cyclic peptides was genuine, the Pgp-Glo$^{TM}$ assay was repeated using linear peptide IPAPIPFP (**32**). Peptide **32** can be thought of a functionless cyanobactin analogue (free N and C-termini, absence of azol(in)e heterocycles, all L-amino acids) and so should act as a negative control. The ATPase activity observed

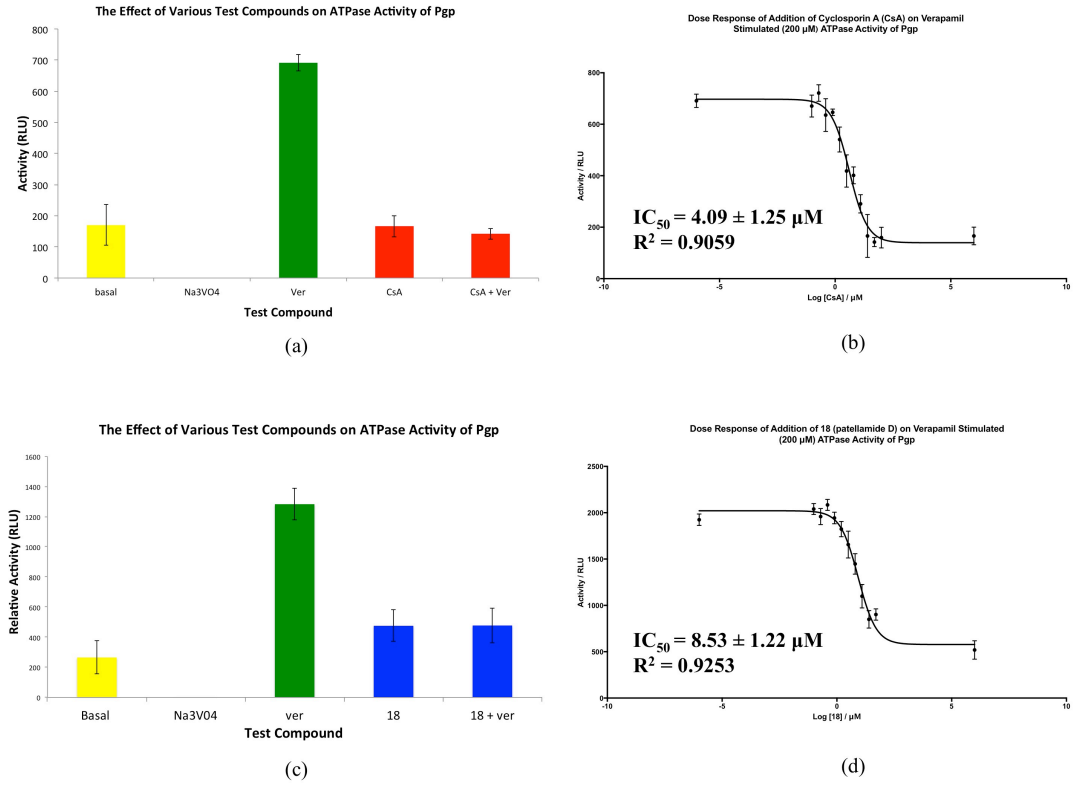



FIGURE 5.19: **Pgp-Glo$^{TM}$ Assay validation** (a) The effect of CsA on ATPase activity of Pgp; basal activity (untreated Pgp) is shown in yellow; 200 µM verapamil treated Pgp (ver) is shown in green; 50 µM CsA and both 50 µM CsA and 200 µM verapamil treated Pgp are shown in red. (b) Dose response of the addition of CsA on 200 µM verapamil stimulated ATPase activity of Pgp. (c) The effect of **18** on ATPase activity of Pgp; basal activity (untreated Pgp) is shown in yellow; 200 µM verapamil treated Pgp (ver) is shown in green; 50 µM **18** and both 50 µM **18** and 200 µM verapamil treated Pgp are shown in blue. (d) Dose response of the addition of **18** on 200 µM verapamil stimulated ATPase activity of Pgp. Error bars represent +/- 1 standard error.

for the addition of 50 µM **32** is, within error, the same as the basal level (Fig. 5.20). Similarly, upon addition of 50 µM **32** to a reaction containing 200 µM verpamil, the observed activity was, within error, the same as that of verapamil alone (Fig. 5.20). The drastically different response of **32** compared with **18** and CsA, is encouraging, and suggests the assay will be suitable to identify a range of responses for structurally varied compounds.

Compounds **1**, **4**, **9**, **10**, and **15-31** were screened for their ability to inhibit the ATPase activity of Pgp stimulated by 200 µM verapamil at 50, 5 and 0.5 µM. Reactions were set up in triplicate over three, ninety-six well plates. In addition, each ninety-six well plate contained the following control reactions in triplicate (nine replicates in total): (1)

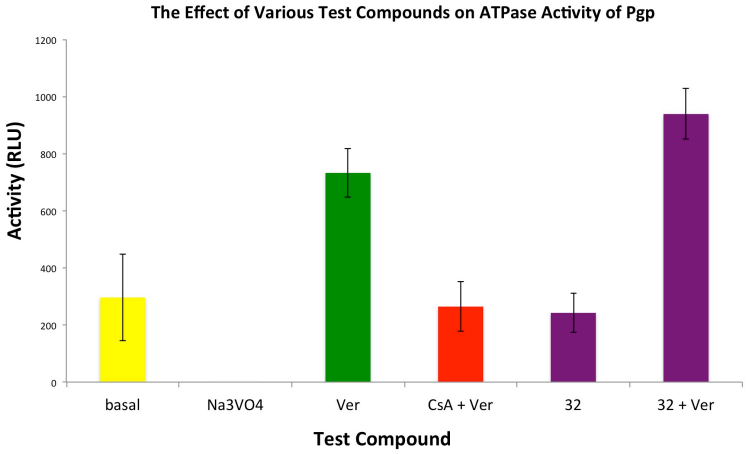
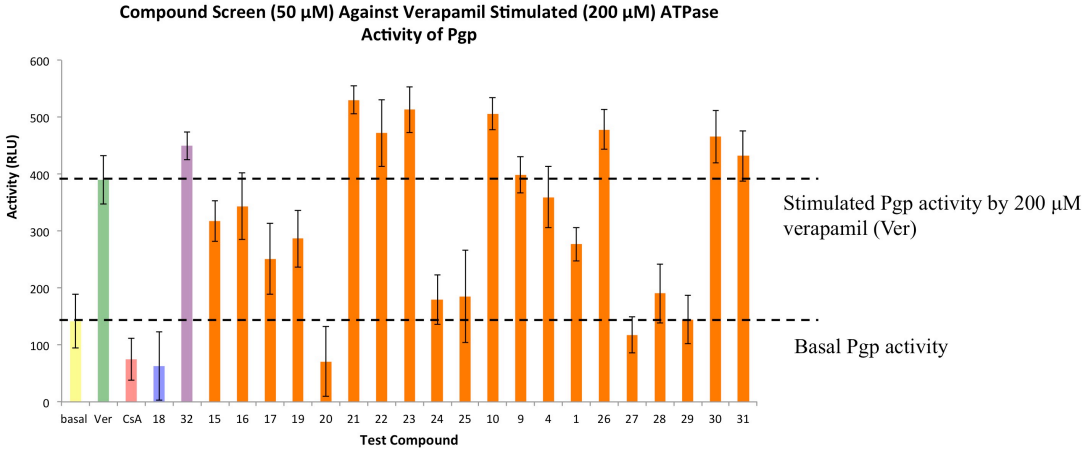


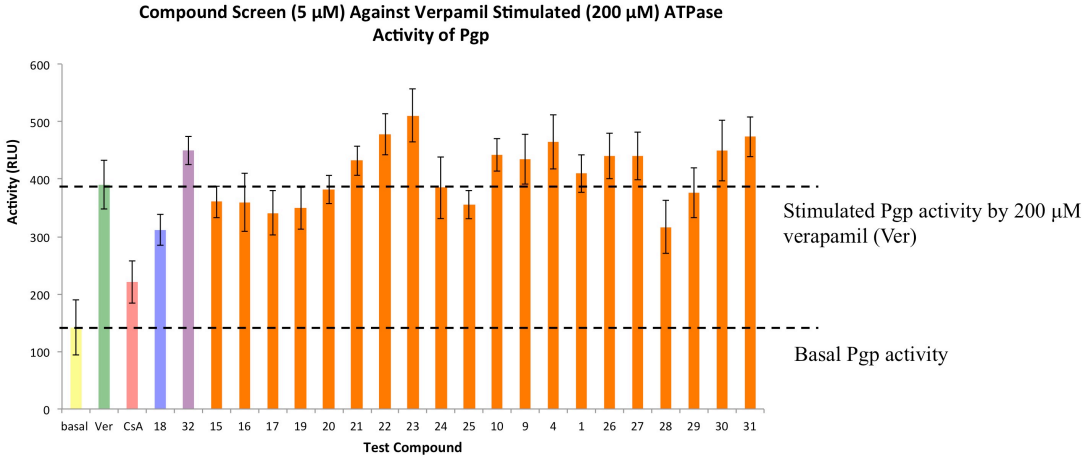

FIGURE 5.20: **Linear peptide 32 as a negative control** The effect of **32** on ATPase activity of Pgp; basal activity (untreated Pgp) is shown in yellow; 200 μM verapamil treated Pgp (ver) is shown in green; 50 μM CsA and 200 μM verapamil treated Pgp is shown in red; and 50 μM **32** and both 50 μM CsA and 200 μM verapamil treated Pgp are shown in purple. Error bars represent +/- 1 standard error.

Pgp only, (2) Pgp + 100 μM $Na_3VO_4$, (3) Pgp + 200 μM verapamil, (4) Pgp + 200 μM verapamil + 50 μM CsA, (5) Pgp + 200 μM verapamil + 5 μM CsA, (6) Pgp + 200 μM verapamil + 0.5 μM CsA, (7) Pgp + 200 μM verapamil + 50 μM **32**. The relative ATPase activity of Pgp in the presence of 200 μM verapamil and 50, 5 and 0.5 μM of each test compound and for the control reactions can be seen in Fig. 5.21. Note control reaction (2) (Pgp + 100 μM $Na_3VO_4$) is not included, as the activity is always 0 RLU by default.

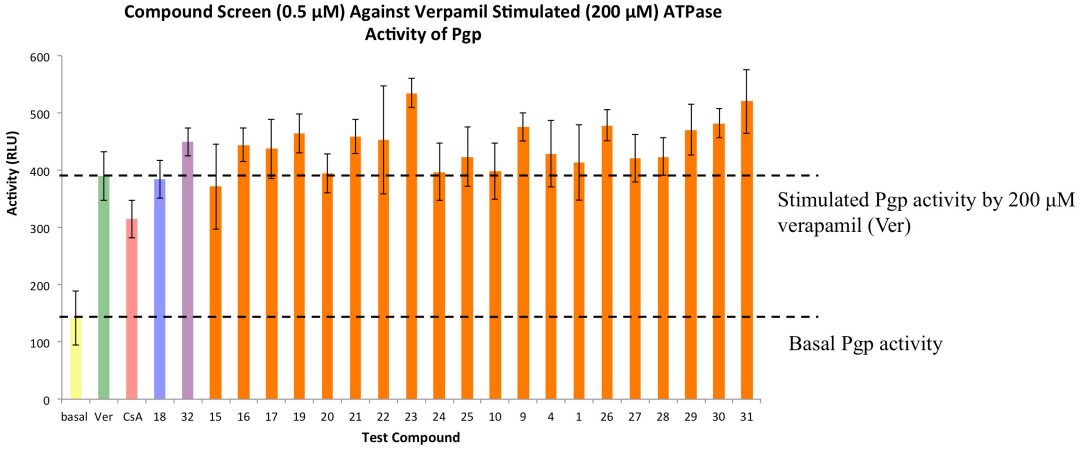
These data revealed that the structurally varied compounds **1**, **4**, **9**, **10**, and **15-31** display a range of activities towards the ATPase activity of Pgp. The range in activity is most noticeable when the compounds are screened at 50 μM, but it is still evident at 5 μM. Compounds **20**, **24**, **25**, **28**, and **29** were identified as hit compounds (Fig. 5.22). Looking at the structures of these compounds, some similarities become immediately obvious. All compounds contain at least three heterocycles (azol(in)e or proline) and the intervening residues (positions 1, 3, 5 and 7) correspond to the native sequence of patellamide D.




(a)




(b)

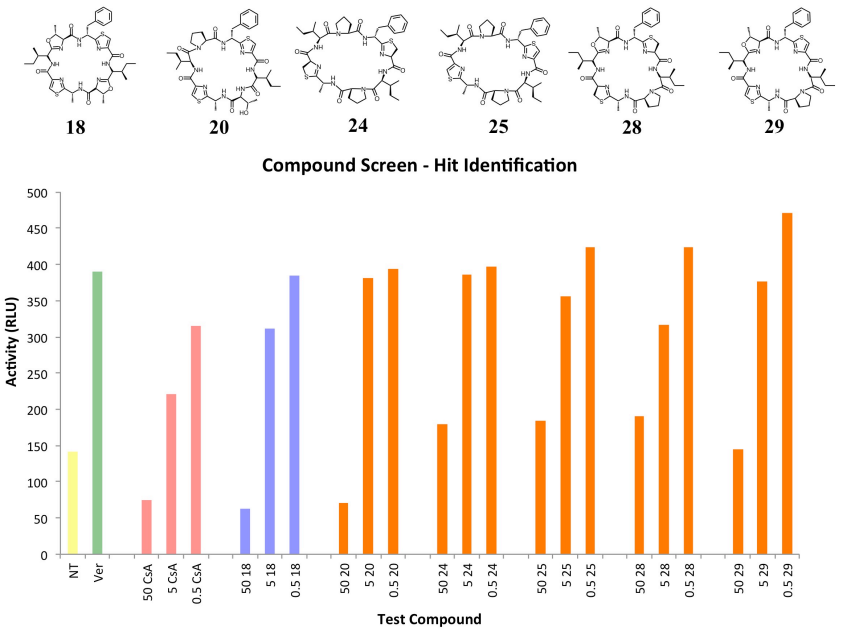



(c)

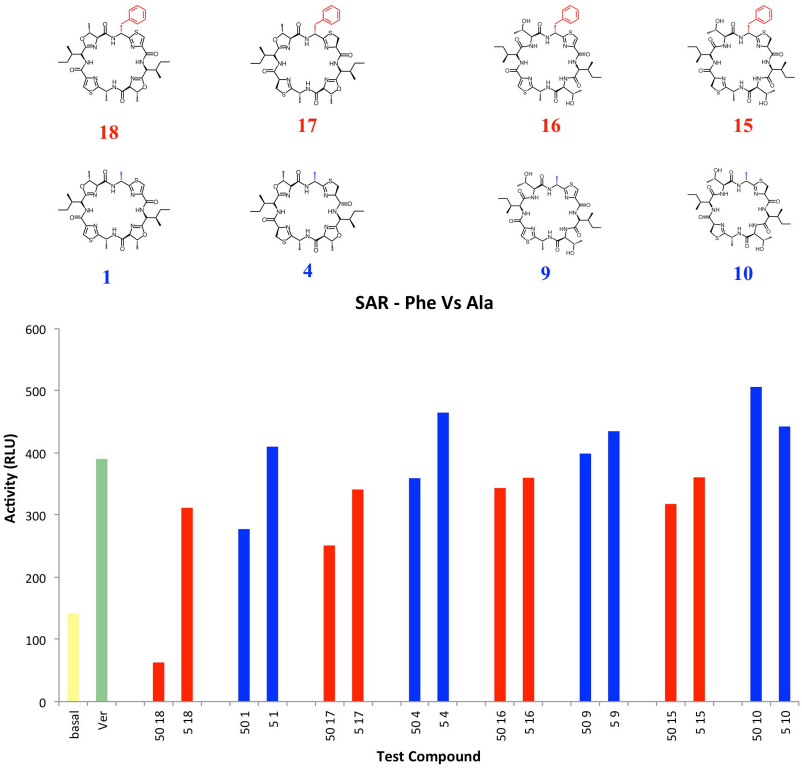
FIGURE 5.21: **SAR compound screen** The effect of compounds **1**, **4**, **9**, **10**, and **15-31**, CsA and **32** on ATPase activity of Pgp stimulated with 200 µM verapamil at (a) 50 µM, (b) 5 µM and (c) 0.5 µM. Basal activity (untreated Pgp) is shown in yellow; 200 µM verapamil-treated Pgp (ver) is shown in green; CsA and 200 µM verapamil-treated Pgp is shown in red; **18** and 200 µM verapamil treated Pgp is shown in blue; 50 µM **32** treated and 200 µM verapamil treated Pgp is shown in purple (note **32** was only screened at 50 µM); Pgp treated with 200 µM verapamil and test compounds **1**, **4**, **9**, **10**, and **15-31** (except **18**) are shown in orange. The upper and lower dashed lines highlight 200 µM verapamil stimulated ATPase activity and basal ATPase activity respectively. Error bars represent +/- 1 standard error.

By looking closely at the Pgp-Glo$^{TM}$ assay data, and by making pairwise comparisons (when available) it is possible to build up a set of structural-activity relationships that explain these observations (Fig. 5.22). *(1) The importance of Phe Vs Ala* (Fig. 5.22): The starting peptides ITACITFCAYD and ITACITACAYD give rise to a family of compounds **15**, **16**, **17**, **18** and **10**, **9**, **4**, **1** respectively, which differ within their respective pair by only a Phe/Ala substitution at position 7. In all examples, exchanging Phe at position 7 with Ala corresponds to a reduction in efficacy, indicating Phe is important for activity. *(2) The importance of the number of heterocycles* (Fig. 5.22): By comparing a subset of the compounds (**18**, **20**, **25**, **28)** which contain three or more heterocycles (azol(in)e or proline) with a subset of compounds containing only two heterocycles (**16**, **23**, **10**, **26)**, it becomes apparent that the compounds with the greater number of heterocycles are significantly more active at both 50 and 5 μM. Furthermore substitution of oxazoline with proline retains good inhibitory strength, suggesting the conformational consequence of the five-membered ring is more important for an interaction with Pgp than the electrochemical properties of the oxazoline ring. These observations are consistant with the identification of compounds **20**, **24**, **25**, **28**, and **29** as hit compounds. *(3) The importance of Thz vs ThH* (Fig. 5.22): Throughout the data there is a general trend that compounds containing the oxidised Thz heterocycles confer better inhibition than their counterparts with reduced ThH heterocycles. This trend is epitomised by the examples in Fig. 5.22, particularly when the compounds are screened at 50 μM. Although this trend does not apply to all the compound pairs within the screen, it is followed more often than not, and so is likely an important factor in activity.

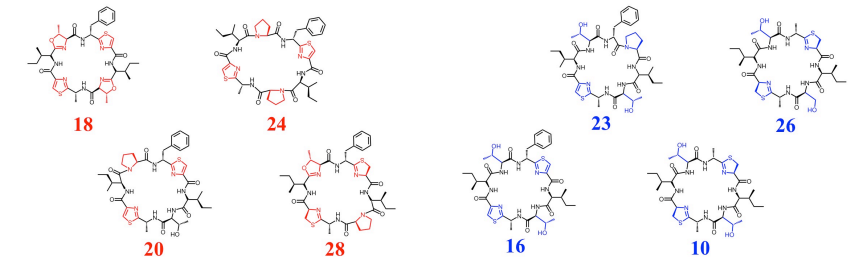
The dose-dependent inhibition of Pgp by the three oxidised hit compounds **20**, **25** and **29** were measured and gave $IC_{50}$ = 6.62, 1.68 and 5.08 μM respectively (Fig. 5.23). These data reinforce the idea that oxazoline can be adequately substituted with proline, and in these cases actually result in more efficacious compounds, compared with the natural compound **18** ($IC_{50}$ = 8.53). In fact the most potent of the three compounds, compound **25**, has both oxazoline residues substituted with proline residues. The least active of these three, compound **20**, contains an unprocessed threonine residue, while in the more active **29**, the remaining threonine residue has been processed to oxazoline.

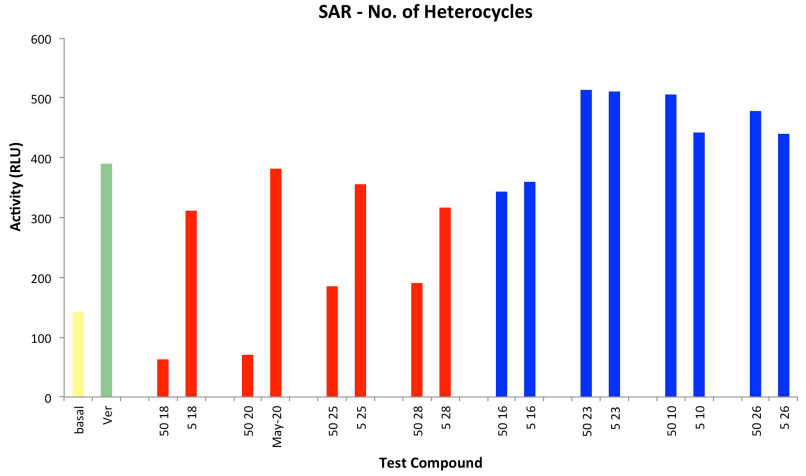

18
20
24
25
28
29
Compound Screen - Hit Identification
Activity (RLU)
0
50
100
150
200
250
300
350
400
450
500
NT
Ver
50 CsA
5 CsA
0.5 CsA
50 18
5 18
0.5 18
50 20
5 20
0.5 20
50 24
5 24
0.5 24
50 25
5 25
0.5 25
50 28
5 28
0.5 28
50 29
5 29
0.5 29
Test Compound

(a)

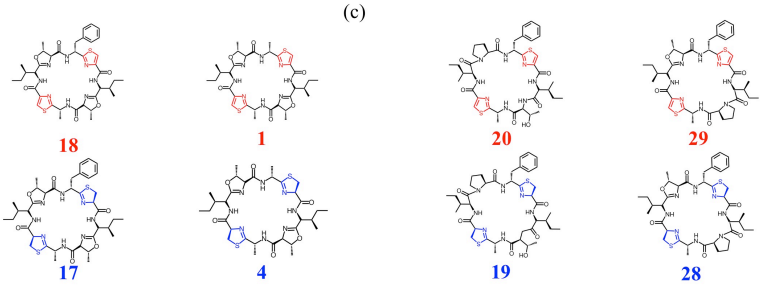

18
17
16
15
1
4
9
10
SAR - Phe Vs Ala
Activity (RLU)
0
100
200
300
400
500
600
basal
Ver
50 18
5 18
50 1
5 1
50 17
5 17
50 4
5 4
50 16
5 16
50 9
5 9
50 15
5 15
50 10
5 10
Test Compound

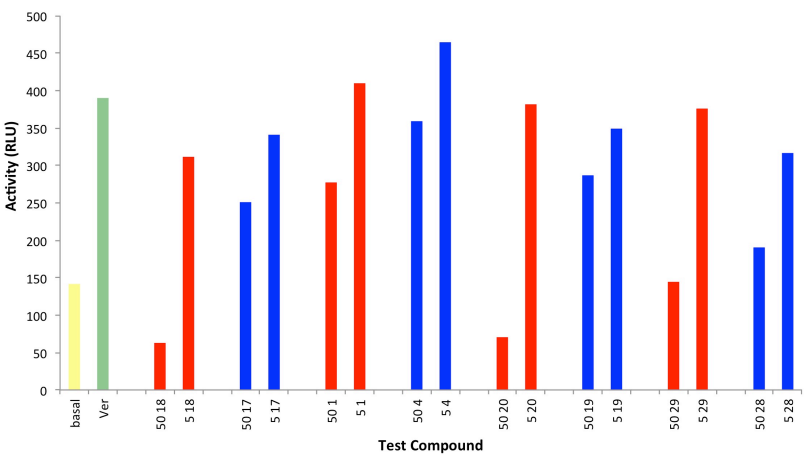
(b)




(c)

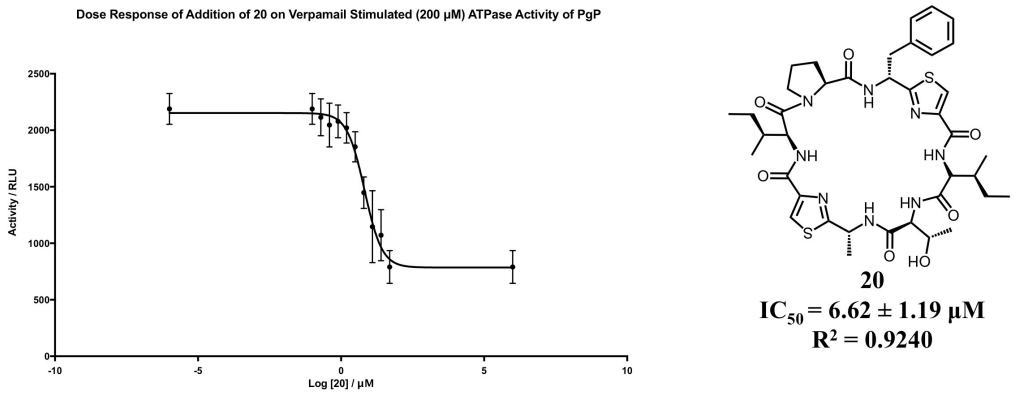



(d)

FIGURE 5.22: **SAR analysis** (a) Identification of compounds **20**, **24**, **25**, **28**, and **29** as hit compounds at 50, 5 and 0.5 µM shown in orange. Basal activity (untreated Pgp) is shown in yellow; 200 µM verapamil treated Pgp (ver) is shown in green; 50, 5 and 0.5 µM CsA and 200 µM verapamil treated Pgp are shown in red; 50, 5 and 0.5 µM **18** and 200 µM verapamil treated Pgp are shown in blue. (b) Activity of Phe-containing compounds compounds **18**, **17**, **16**, **15** (red) vs Ala-containing compounds **1**, **4**, **9**, **10** (blue). (c) Activity of compounds containing three or more heterocycles **18**, **20**, **24**, **28** (red) vs compounds containing two heterocycles **16**, **23**, **10**, **26** (blue). (d) Activity of Thz containing compounds **18**, **1**, **20**, **29** (red) vs ThH containing compounds **17**, **4**, **19**, **28** (blue). Basal activity (untreated Pgp) is shown in yellow; 200 µM verapamil treated Pgp (ver) is shown in green.

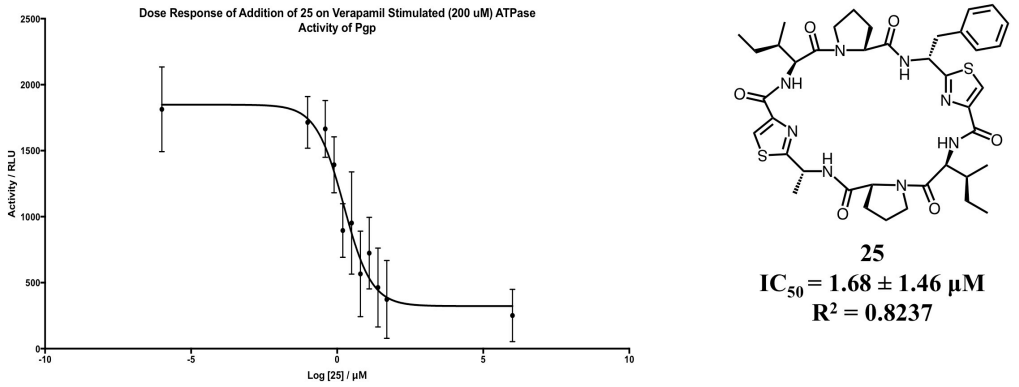
From these data alone, further SAR rules can be proposed. (1) Proline is more efficacious than oxazoline (**25** is the best compound tested, and all three compounds are more active than **18**), and (2) four heterocycles are more potent than three heterocycles (**29** >**20**). However the effect of rule (1) is more significant than that of rule (2) (**20** >**18**). Further work is required to test these hypotheses and to investigate these SARs more thoroughly.






**20**
**$IC_{50}$ = 6.62 ± 1.19 µM**
**$R^2$ = 0.9240**

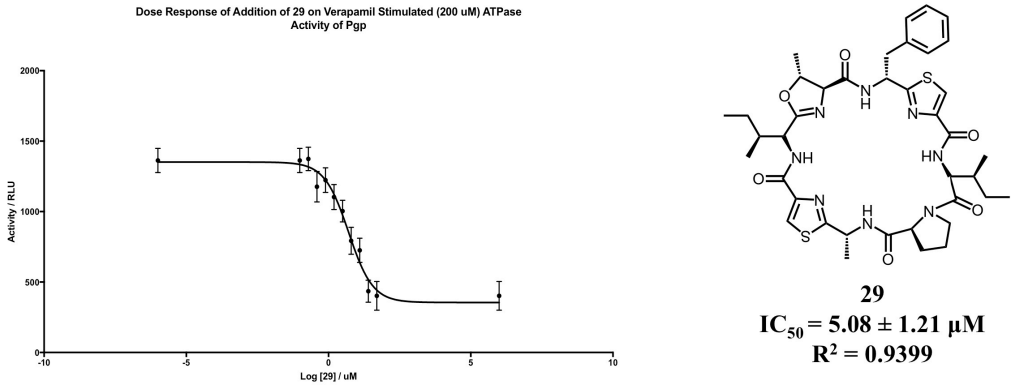
(a)






**25**
**$IC_{50}$ = 1.68 ± 1.46 µM**
**$R^2$ = 0.8237**

(b)






**29**
**$IC_{50}$ = 5.08 ± 1.21 µM**
**$R^2$ = 0.9399**

(c)

FIGURE 5.23: **Dose-dependant Pgp inhibition of hit compounds** Dose response of addition of (a) **20**, (b) **25**, (c) **29** on 200 µM verapamil stimulated ATPase activity of Pgp. Error bars represent +/- 1 standard error.

## 5.4 Conclusions

The one-pot *in vitro* synthesis provides a route to the production of natural and unnatural, highly modified cyclic peptides, which is amenable to both scale and diversity. The modified one-pot method maintains the flexibility of the original *in vitro* synthesis[47], however it does not require multiple purification steps, making it much faster and more efficient. Additionally, as the need for trypsin proteolysis has been dispensed with, the one-pot *in vitro* synthesis can, in principle incorporate lysine and arginine residues into the final products. A major strength of the *in vitro* synthesis is the a ability to produce multiple products from a single starting peptide, simply by changing the combination of enzymes or altering the reaction conditions. Using the biotechnological toolkit described in this thesis, it is currently possible to produce up to five products from a single precursor peptide (Fig. 5.24). This toolkit, and thus product diversity can be expanded by incorporating other, cyanobactin PTM enzymes, including various prenyl-transferase enzymes[57,59], and potentially PTM enzymes from other RiPPs.

The compatibility of the one-pot *in vitro* synthesis with synthetic substrates combines the efficiency and the economical and environmental advantages of biotechnology with the flexibility of chemical synthesis, and can be used to introduce diversity into peptidic macrocycle scaffolds beyond amino acids[58]. The ease of preparation of the one-pot method make it suitable for the manufacture of small, focused libraries, based on a known starting point; as demonstrated by the synthesis of greater than twenty analogues of patellamide D. Other method developments detailed in this chapter also contribute to the overall efficiency of cyclic peptide production and analysis. The developed extraction procedure using n-BuOH selectively isolates peptide macrocycles from complex mixtures of enzymes, co-factors and linear peptides. It provides a straightforward means to access purer product, which is potentially applicable on a multi-well format, and so compatible with library production. Additionally the validated-qNMR methodology provides a universally applicable method to quantify synthesis products which lack a fluorophore, and so cannot be accurately quantified by other means.

By synthesising over twenty analogues of patellamide D via the one-pot *in vitro* synthesis, and testing them using the Pgp-Glo$^{TM}$ assay for their ability to inhibit the ATPase activity of Pgp, it has been possible to assign some preliminary structural-activity

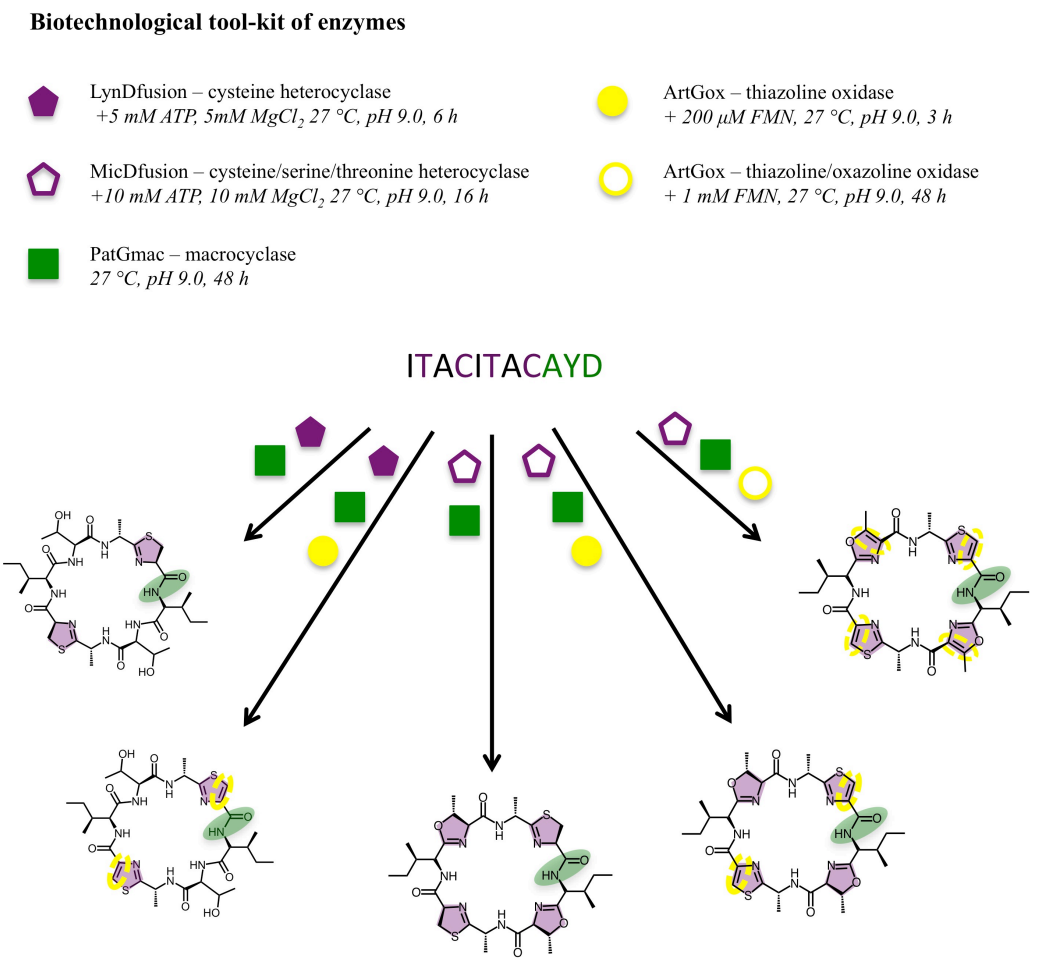



FIGURE 5.24: **One-pot *in vitro* synthesis products** Using the current biotechnological toolkit described in this thesis it is possible to produce up to five products from a single starting peptide.

relationships for Pgp inhibition. In summary: substituting phenylalanine at position 7 with alanine greatly reduced efficacy, and compounds with more than two heterocycles (azol(in)e or proline) confer the best activities. Interestingly oxazoline moieties can be substituted with proline residues without loss of activity, and thiazole heterocycles are generally more active than thiazoline. Moreover three analogues, containing a mixture of prolines and thiazoles, more potent than the natural compound have been identified.

## 5.5 Future Work

Attempts to optimise the one-pot *in vitro* synthesis reaction conditions are ongoing, with the need to reduce the amount of enzyme required to accomplish each biotransformation

of paramount importance when operating on a larger scale. One possible solution is to immobilise the enzymes onto beads. This would not only simplify downstream purification of the products, but potentially enable the enzymes to be recycled, reducing the need to regularly purify large amounts of protein.

Work is underway to combine the one-pot *in vitro* synthesis and solid-phase peptide synthesis, ultimately on the one bead one compound scale. This would enable a route to larger library design, and potentially the identification of new, biologically active peptide macrocycles.

Further work is required to test the SAR hypotheses proposed in this chapter, and the work can be expanded to test more compounds containing a greater range of amino acids, to more thoroughly assess SAR of Pgp inhibition. Additionally alternate functionality can be tested, including oxazole rings and prenylated residues. Ultimately bespoke macrocycles with increased efficacy can be designed, synthesised and tested.

# Bibliography


[1] J. Hong and H. Luesch. Largazole: From discovery to broad-spectrum therapy. *Natural Product Reports*, 29(4):449–456, April 2012. ISSN 1460-4752. doi: 10.1039/C2NP00066K.

[2] K. Yamada, M. Kodaira, S. Shinoda, K. Komagoe, H. Oku, R. Katakai, T. Katsu, and I. Matsuo. Structure?activity relationships of gramicidin S analogs containing (β-3-pyridyl)-α,β-dehydroalanine residues on membrane permeability. *MedChemComm*, 2(7):644–649, July 2011. ISSN 2040-2511. doi: 10.1039/C1MD00081K.

[3] S. Ho, N. Clipstone, L. Timmermann, J. Northrop, I. Graef, D. Fiorentino, J. Nourse, and G. R. Crabtree. The Mechanism of Action of Cyclosporin A and FK506. *Clinical Immunology and Immunopathology*, 80(3):S40–S45, September 1996. ISSN 0090-1229. doi: 10.1006/clin.1996.0140.

[4] K. Sabol and T. Gumbo. Anidulafungin in the treatment of invasive fungal infections. *Therapeutics and Clinical Risk Management*, 4(1):71–78, February 2008. ISSN 1176-6336.

[5] D. P. Levine. Vancomycin: A History. *Clinical Infectious Diseases*, 42(Supplement 1):S5–S12, January 2006. ISSN 1058-4838, 1537-6591. doi: 10.1086/491709.

[6] X. Yu and D. Sun. Macrocyclic Drugs and Synthetic Methodologies toward Macrocycles. *Molecules*, 18(6):6230–6268, May 2013. doi: 10.3390/molecules18066230.

[7] D. J. Craik, D. P. Fairlie, S. Liras, and D. Price. The future of peptide-based drugs. *Chemical Biology & Drug Design*, 81(1):136–147, January 2013. ISSN 1747-0285. doi: 10.1111/cbdd.12055.

[8] J. M. Mason. Design and development of peptides and peptide mimetics as antagonists for therapeutic intervention. *Future Medicinal Chemistry*, 2(12):1813–1822, December 2010. ISSN 1756-8919, 1756-8927. doi: 10.4155/fmc.10.259.

[9] T. Uhlig, T. Kyprianou, F. G. Martinelli, C. A. Oppici, D. Heiligers, D. Hills, X. R. Calvo, and P. Verhaert. The emergence of peptides in the pharmaceutical business: From exploration to exploitation. *EuPA Open Proteomics*, 4:58–69, September 2014. ISSN 2212-9685. doi: 10.1016/j.euprot.2014.05.003.

[10] M. Gao, K. Cheng, and H. Yin. Targeting protein-protein interfaces using macrocyclic peptides. *Biopolymers*, 104(4):310–316, February 2015. ISSN 0006-3525. doi: 10.1002/bip.22625.

[11] A. Roxin and G. Zheng. Flexible or fixed: a comparative review of linear and cyclic cancer-targeting peptides. *Future Medicinal Chemistry*, 4(12):1601–1618, August 2012. ISSN 1756-8927. doi: 10.4155/fmc.12.75.

[12] G. Zinzalla and D. E. Thurston. Targeting protein-protein interactions for therapeutic intervention: a challenge for the future. *Future Medicinal Chemistry*, 1(1):65–93, April 2009. ISSN 1756-8919. doi: 10.4155/fmc.09.12.

[13] A. T. Bockus, C. M. McEwen, and S. R. Lokey. Form and function in cyclic peptide natural products: a pharmacokinetic perspective. *Current Topics in Medicinal Chemistry*, 13(7):821–836, 2013. ISSN 1873-4294.

[14] S. Namjoshi and H. A. E. Benson. Cyclic peptides as potential therapeutic agents for skin disorders. *Biopolymers*, 94(5):673–680, 2010. ISSN 0006-3525. doi: 10.1002/bip.21476.

[15] W. M. Hewitt, S. S. F. Leung, C. R. Pye, A. R. Ponkey, M. Bednarek, M. P. Jacobson, and S. R. Lokey. Cell-permeable cyclic peptides from synthetic libraries inspired by natural products. *Journal of the American Chemical Society*, 137(2):715–721, January 2015. ISSN 1520-5126. doi: 10.1021/ja508766b.

[16] W. E. Houssen and M. Jaspars. Azole-Based Cyclic Peptides from the Sea Squirt Lissoclinum Patella: Old Scaffolds, New Avenues. *ChemBioChem*, 11(13):1803–1815, 2010. ISSN 1439-7633. doi: 10.1002/cbic.201000230.

[17] C. J. White and A. K. Yudin. Contemporary strategies for peptide macrocyclization. *Nature Chemistry*, 3(7):509–524, July 2011. ISSN 1755-4330. doi: 10.1038/nchem.1062.

[18] R. Finking and M. A. Marahiel. Biosynthesis of nonribosomal peptides1. *Annual review of microbiology*, 58:453–488, 2004. ISSN 0066-4227. doi: 10.1146/annurev.micro.58.030603.123615.

[19] Noah A. Bindman and Wilfred A. Van Der Donk. RiPPs: Ribosomally Synthesized and Posttranslationally Modified Peptides. In A. Osbourn, R. J. Goss, and G. T. Carter, editors, *Natural Products*, pages 195–217. John Wiley & Sons, Inc., 2014. ISBN 9781118794623.

[20] P. G. Arnison and *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural product reports*, 30(1):108–160, January 2013. ISSN 0265-0568. doi: 10.1039/c2np20085f.

[21] K. Sivonen, N. Leikoski, D. P. Fewer, and J. Jokela. Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. *Applied microbiology and biotechnology*, 86 (5):1213–1225, May 2010. ISSN 1432-0614. doi: 10.1007/s00253-010-2482-x.

[22] X. Yang and W. A. van der Donk. Ribosomally synthesized and post-translationally modified peptide natural products: new insights into the role of leader and core peptides during biosynthesis. *Chemistry (Weinheim an der Bergstrasse, Germany)*, 19(24):7662–7677, June 2013. ISSN 1521-3765. doi: 10.1002/chem.201300401.

[23] D. E. Williams, R. E. Moore, and V. J. Paul. The structure of ulithiacyclamide B. Antitumor evaluation of cyclic peptides and macrolides from Lissoclinum patella. *Journal of Natural Products*, 52(4):732–739, August 1989. ISSN 0163-3864.

[24] X. Fu, T. Do, F. J. Schmitz, V. Andrusevich, and M. H. Engel. New Cyclic Peptides from the Ascidian Lissoclinum patella. *Journal of Natural Products*, 61 (12):1547–1551, December 1998. ISSN 0163-3864. doi: 10.1021/np9802872.

[25] A. B. Williams and R. S. Jacobs. A marine natural product, patellamide D, reverses multidrug resistance in a human leukemic cell line. *Cancer letters*, 71 (1-3):97–102, 1993. ISSN 0304-3835. doi: 10.1016/0304-3835(93)90103-G.

[26] J. Martins and V. Vasconcelos. Cyanobactins from Cyanobacteria: Current Genetic and Chemical State of Knowledge. *Marine Drugs*, 13(11):6910–6946, November 2015. doi: 10.3390/md13116910.

[27] J. Koehnke, A. F. Bent, W. E. Houssen, G. Mann, M. Jaspars, and J. H. Naismith. The structural biology of patellamide biosynthesis. *Current Opinion in Structural Biology*, 29:112–121, December 2014. ISSN 1879-033X. doi: 10.1016/j.sbi.2014.10.006.

[28] E. W. Schmidt, J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, Margo G. Haygood, and J. Ravel. Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20):7315–7320, May 2005. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0501424102.

[29] G. J. Schuurhuis, H. J. Broxterman, G. J. Ossenkoppele, J. P. Baak, C. A. Eekman, C. M. Kuiper, N. Feller, T. H. van Heijningen, E. Klumper, and R. Pieters. Functional multidrug resistance phenotype associated with combined overexpression of Pgp/MDR1 and MRP together with 1-beta-D-arabinofuranosylcytosine sensitivity may predict clinical response in acute myeloid leukemia. *Clinical cancer research: an official journal of the American Association for Cancer Research*, 1(1):81–93, January 1995. ISSN 1078-0432.

[30] J. H. Lin and M. Yamazaki. Role of P-glycoprotein in pharmacokinetics: clinical implications. *Clinical pharmacokinetics*, 42(1):59–98, 2003. ISSN 0312-5963.

[31] B. Ryffel, G. Woerly, C. Rodriguez, and B. M. Foxwell. Identification of the multidrug resistance-related membrane glycoprotein as an acceptor for cyclosporine. *Journal of receptor research*, 11(1-4):675–686, 1991. ISSN 0197-5110.

[32] S. G. Aller, J. Yu, A. Ward, Y. Weng, S. Chittaboina, R. Zhuo, P. M. Harrell, Y. T. Trinh, Q. Zhang, I. L. Urbatsch, and G. Chang. Structure of P-Glycoprotein Reveals a Molecular Basis for Poly-Specific Drug Binding. *Science*, 323(5922): 1718–1722, March 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1168750.

[33] L. Huo, S. Rachid, M. Stadler, S.C. Wenzel, and R. Mller. Synthetic Biotechnology to Study and Engineer Ribosomal Bottromycin Biosynthesis. *Chemistry & Biology*, 19(10):1278–1287, October 2012. ISSN 1074-5521. doi: 10.1016/j.chembiol.2012.08.013.

[34] W. J. K. Crone, F. J. Leeper, and A. W. Truman. Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. *Chemical Science*, 3 (12):3516–3521, October 2012. ISSN 2041-6539. doi: 10.1039/C2SC21190D.

[35] T. J. Oman and W. A. van der Donk. Follow the leader: The use of leader peptides to guide natural product biosynthesis. *Nature chemical biology*, 6(1):9–18, January 2010. ISSN 1552-4450. doi: 10.1038/nchembio.286.

[36] J. Lee, J. McIntosh, B. J. Hathaway, and E. W. Schmidt. Using Marine Natural Products to Discover a Protease that Catalyzes Peptide Macrocyclization of Diverse Substrates. *Journal of the American Chemical Society*, 131(6):2122–2124, February 2009. ISSN 0002-7863. doi: 10.1021/ja8092168.

[37] E. W. Schmidt. The hidden diversity of ribosomal peptide natural products. *BMC biology*, 8(83):1–4, 2010. ISSN 1741-7007. doi: 10.1186/1741-7007-8-83.

[38] M. S. Donia, J. Ravel, and E. W. Schmidt. A global assembly line to cyanobactins. *Nature chemical biology*, 4(6):341–343, June 2008. ISSN 1552-4450. doi: 10.1038/nchembio.84.

[39] T. Katoh, Y. Goto, S. M. Reza, and H. Suga. Ribosomal synthesis of backbone macrocyclic peptides. *Chemical Communications*, 47(36):9946–9958, 2011. ISSN 1359-7345, 1364-548X. doi: 10.1039/c1cc12647d.

[40] J. A. McIntosh, M. S. Donia, and E. W. Schmidt. Insights into Heterocyclization from Two Highly Similar Enzymes. *Journal of the American Chemical Society*, 132(12):4089–4091, March 2010. ISSN 0002-7863. doi: 10.1021/ja9107116.

[41] Y. Goto, Y. Ito, Y. Kato, S. Tsunoda, and H. Suga. One-pot synthesis of azoline-containing peptides in a cell-free translation system integrated with a posttranslational cyclodehydratase. *Chemistry & Biology*, 21(6):766–774, June 2014. ISSN 1879-1301. doi: 10.1016/j.chembiol.2014.04.008.

[42] W. E. Houssen, J. Koehnke, D. Zollman, J. Vendome, A. Raab, M. C. M. Smith, J. H. Naismith, and M. Jaspars. The discovery of new cyanobactins from Cyanothece PCC 7425 defines a new signature for processing of patellamides. *Chembiochem: a European journal of chemical biology*, 13(18):2683–2689, December 2012. ISSN 1439-7633. doi: 10.1002/cbic.201200661.

[43] J. Koehnke, A. Bent, W. E. Houssen, D. Zollman, F. Morawitz, S. Shirran, J. Vendome, A. F. Nneoyiegbe, L. Trembleau, C. H. Botting, M. C. M. Smith, M. Jaspars, and J. H. Naismith. The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. *Nature structural & molecular biology*, 19(8):767–772, August 2012. ISSN 1545-9985. doi: 10.1038/nsmb.2340.

[44] V. Agarwal, E. Pierce, J. McIntosh, E. W. Schmidt, and S. K. Nair. Structures of cyanobactin maturation enzymes define a family of transamidating proteases. *Chemistry & biology*, 19(11):1411–1422, November 2012. ISSN 1879-1301. doi: 10.1016/j.chembiol.2012.09.012.

[45] B. F. Milne, P. F. Long, A. Starcevic, D. Hranueli, and M. Jaspars. Spontaneity in the patellamide biosynthetic pathway. *Organic & Biomolecular Chemistry*, 4 (4):631–638, February 2006. ISSN 1477-0539. doi: 10.1039/B515938E.

[46] M. S. Donia, B. J. Hathaway, S. Sudek, M. G. Haygood, M. J. Rosovitz, J. Ravel, and E. W. Schmidt. Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nature Chemical Biology*, 2(12):729–735, December 2006. ISSN 1552-4450. doi: 10.1038/nchembio829.

[47] W. E. Houssen, A. F. Bent, A. R. McEwan, N. Pieiller, J. Tabudravu, J. Koehnke, G. Mann, R. I. Adaba, L. Thomas, U. W. Hawas, H. Liu, U. Schwarz-Linek, M. C. M. Smith, J. H. Naismith, and M. Jaspars. An Efficient Method for the In Vitro Production of Azol(in)e-Based Cyclic Peptides. *Angewandte Chemie (International Ed. in English)*, 53(51):14171–14174, December 2014. ISSN 1433-7851. doi: 10.1002/anie.201408082.

[48] J. A. McIntosh and E. W. Schmidt. Marine Molecular Machines: Heterocyclization in Cyanobactin Biosynthesis. *ChemBioChem*, 11(10):1413–1421, 2010. ISSN 1439-7633. doi: 10.1002/cbic.201000196.

[49] K. L. Dunbar, J. O. Melby, and D. A. Mitchell. YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nature chemical biology*, 8(6): 569–575, June 2012. ISSN 1552-4469. doi: 10.1038/nchembio.944.

[50] K. L. Dunbar and D. A. Mitchell. Insights into the mechanism of peptide cyclodehydrations achieved through the chemoenzymatic generation of amide derivatives. *Journal of the American Chemical Society*, 135(23):8692–8701, June 2013. ISSN 1520-5126. doi: 10.1021/ja4029507.

[51] J. Koehnke, A. F. Bent, D. Zollman, K. Smith, W. E. Houssen, X. Zhu, G. Mann, T. Lebl, R. Scharff, S. Shirran, C. H. Botting, M. Jaspars, U. Schwarz-Linek, and J. H. Naismith. The cyanobactin heterocyclase enzyme: a processive adenylase that operates with a defined order of reaction. *Angewandte Chemie (International Ed. in English)*, 52(52):13991–13996, December 2013. ISSN 1521-3773. doi: 10.1002/anie.201306302.

[52] D. Sardar, E. Pierce, J. A. McIntosh, and E. W. Schmidt. Recognition Sequences and Substrate Evolution in Cyanobactin Biosynthesis. *ACS Synthetic Biology*, 4 (2):167–176, February 2015. doi: 10.1021/sb500019b.

[53] R. J. Siezen and J. A. Leunissen. Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Science : A Publication of the Protein Society*, 6(3): 501–523, March 1997. ISSN 0961-8368.

[54] M. D. Tianero, E. Pierce, S. Raghuraman, D. Sardar, J. A. McIntosh, J. R. Heemstra, Z. Schonrock, B. C. Covington, J. A. Maschek, J. E. Cox, B. O. Bachmann, B. M. Olivera, D. E. Ruffner, and E. W. Schmidt. Metabolic model for diversity-generating biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 113(7):1772–1777, February 2016. ISSN 1091-6490. doi: 10.1073/pnas.1525438113.

[55] J. A. McIntosh, C. R. Robertson, V. Agarwal, S. K. Nair, G. W. Bulaj, and E. W. Schmidt. Circular Logic: Nonribosomal Peptide-like Macrocyclization with a Ribosomal Peptide Catalyst. *Journal of the American Chemical Society*, 132 (44):15499–15501, November 2010. ISSN 0002-7863. doi: 10.1021/ja1067806.

[56] D. E. Ruffner, E. W. Schmidt, and J. R. Heemstra. Assessing the combinatorial potential of the RiPP cyanobactin tru pathway. *ACS synthetic biology*, 4(4):482–492, April 2015. ISSN 2161-5063. doi: 10.1021/sb500267d.

[57] M. D. B Tianero, M. S. Donia, T. S. Young, P. G. Schultz, and E. W. Schmidt. Ribosomal route to small-molecule diversity. *Journal of the American Chemical Society*, 134(1):418–425, January 2012. ISSN 1520-5126. doi: 10.1021/ja208278k.

[58] E. Oueis, M. Jaspars, N. J. Westwood, and J. H. Naismith. Enzymatic Macrocyclization of 1,2,3-Triazole Peptide Mimetics. *Angewandte Chemie International Edition*, 55(19):5842–5845, May 2016. ISSN 1521-3773. doi: 10.1002/anie.201601564.

[59] J. A. McIntosh, M. S. Donia, S. K. Nair, and E. W. Schmidt. Enzymatic basis of ribosomal peptide prenylation in cyanobacteria. *Journal of the American Chemical Society*, 133(34):13698–13705, August 2011. ISSN 0002-7863. doi: 10.1021/ja205458h.

[60] A. F. Bent, J. Koehnke, W. E. Houssen, M. C. M. Smith, M. Jaspars, and J. H. Naismith. Structure of PatF from Prochloron didemni. *Acta crystallographica. Section F, Structural biology and crystallization communications*, 69(6):618–623, June 2013. ISSN 1744-3091. doi: 10.1107/S1744309113012931.

[61] F. J. Schmitz, M. B. Ksebati, J. S. Chang, J. L. Wang, B. M. Hossain, D. Van der Helm, M. H. Engel, A. Serban, and J. A. Silfer. Cyclic peptides from the ascidian Lissoclinum patella: conformational analysis of patellamide D by x-ray analysis and molecular modeling. *The Journal of Organic Chemistry*, 54(14):3463–3472, July 1989. ISSN 0022-3263. doi: 10.1021/jo00275a036.

[62] P. Wipf and Y. Uto. Total synthesis and revision of stereochemistry of the marine metabolite trunkamide A. *The Journal of organic chemistry*, 65(4):1037–1049, February 2000. ISSN 0022-3263.

[63] J. M. Caba, I. M. Rodriguez, I. Manzanares, E. Giralt, and F. Albericio. Solid-phase total synthesis of trunkamide A(1). *The Journal of organic chemistry*, 66 (23):7568–7574, November 2001. ISSN 0022-3263.

[64] N. Ziemert, K. Ishida, P. Quillardet, C. Bouchier, C. Hertweck, N. T. de Marsac, and E. Dittmann. Microcyclamide biosynthesis in two strains of Microcystis

aeruginosa: from structure to genes and vice versa. *Applied and environmental microbiology*, 74(6):1791–1797, March 2008. ISSN 1098-5336. doi: 10.1128/AEM.02392-07.

[65] R. Banker and S. Carmeli. Tenuecyclamides A-D, cyclic hexapeptides from the cyanobacterium Nostoc spongiaeforme var. tenue. *Journal of natural products*, 61 (10):1248–1251, October 1998. ISSN 0163-3864. doi: 10.1021/np980138j.

[66] S. Sudek, M. G. Haygood, D. T. A. Youssef, and E. W Schmidt. Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium Trichodesmium erythraeum, predicted from the genome sequence. *Applied and environmental microbiology*, 72(6):4382–4387, June 2006. ISSN 0099-2240. doi: 10.1128/AEM.00380-06.

[67] X. Salvatella, J. M. Caba, F. Albericio, and E. Giralt. Solution Structure of the Antitumor Candidate Trunkamide A by 2d NMR and Restrained Simulated Annealing Methods. *The Journal of Organic Chemistry*, 68(2):211–215, January 2003. ISSN 0022-3263. doi: 10.1021/jo026464s.

[68] S. T. Allard, M. F. Giraud, and J. H. Naismith. Epimerases: structure, function and mechanism. *Cellular and molecular life sciences: CMLS*, 58(11):1650–1665, October 2001. ISSN 1420-682X.

[69] T. Yoshimura and N. Esak. Amino acid racemases: functions and mechanisms. *Journal of bioscience and bioengineering*, 96(2):103–109, 2003. ISSN 1389-1723.

[70] D. B. Stein, U. Linne, and M. A. Marahiel. Utility of epimerization domains for the redesign of nonribosomal peptide synthetases. *The FEBS journal*, 272(17):4506–4520, September 2005. ISSN 1742-464X. doi: 10.1111/j.1742-4658.2005.04871.x.

[71] K. Hoffmann, E. Schneider-Scherzer, H. Kleinkauf, and R. Zocher. Purification and characterization of eucaryotic alanine racemase acting as key enzyme in cyclosporin biosynthesis. *The Journal of biological chemistry*, 269(17):12710–12714, April 1994. ISSN 0021-9258.

[72] S. D. Heck, P. R. Faraci, W. S .and Kelbaugh, N. A. Saccomano, P. F. Thadeio, and R. A. Volkmann. Posttranslational amino acid epimerization: enzyme-catalyzed isomerization of amino acid residues in peptide chains. *Proceedings of the National*

*Academy of Sciences of the United States of America*, 93(9):4036–4039, April 1996. ISSN 0027-8424.

[73] P. Garca-Reynaga and M. S. VanNieuwenhze. A New Total Synthesis of Patellamide A. *Organic Letters*, 10(20):4621–4623, October 2008. ISSN 1523-7060. doi: 10.1021/ol801895y.

[74] S. You, H. Razavi, and J. W. Kelly. A biomimetic synthesis of thiazolines using hexaphenyloxodiphosphonium trifluoromethanesulfonate. *Angewandte Chemie (International Ed. in English)*, 42(1):83–85, January 2003. ISSN 1433-7851.

[75] U. Schreiber, R. Gademann, P. J. Ralph, and A. W. D. Larkum. Assessment of Photosynthetic Performance of Prochloron in Lissoclinum patella in hospite by Chlorophyll Fluorescence Measurements. *Plant and Cell Physiology*, 38(8):945–951, January 1997. ISSN 0032-0781, 1471-9053.

[76] Y. Nov. When Second Best Is Good Enough: Another Probabilistic Look at Saturation Mutagenesis. *Applied and Environmental Microbiology*, 78(1):258–262, January 2012. ISSN 0099-2240, 1098-5336. doi: 10.1128/AEM.06265-11.

[77] T. S. Young and P. G. Schultz. Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *Journal of Biological Chemistry*, 285(15):11039–11044, April 2010. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.R109.091306.

[78] V. J. J. Martin, D. J. Pitera, S. T. Withers, J. D. Newman, and J. D. Keasling. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology*, 21(7):796–802, July 2003. ISSN 1087-0156. doi: 10.1038/nbt833.

[79] H. Liu and J. H. Naismith. A simple and efficient expression and purification system using two newly constructed vectors. *Protein expression and purification*, 63(2):102–111, February 2009. ISSN 1096-0279. doi: 10.1016/j.pep.2008.09.008.

[80] H. Liu and J. H. Naismith. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*, 8(1):91, December 2008. ISSN 1472-6750. doi: 10.1186/1472-6750-8-91.

[81] J. Jancarik and S. H. Kim. Sparse matrix sampling: a screening method for crystallization of proteins. *Journal of Applied Crystallography*, 24(4):409–411, August 1991. ISSN 00218898. doi: 10.1107/S0021889891004430.

[82] P. D. Adams, P. V. Afonine, G. Bunkczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*, 66(2):213–221, January 2010. ISSN 0907-4449. doi: 10.1107/S0907444909052925.

[83] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography*, 60(12-1):2126–2132, December 2004. ISSN 0907-4449. doi: 10.1107/S0907444904019158.

[84] G. N. Murshudov, P. Skubk, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, and A. A. Vagin. REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography*, 67(4):355–367, April 2011. ISSN 1399-0047. doi: 10.1107/S0907444911001314.

[85] J. Painter and E. A. Merritt. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta crystallographica. Section D, Biological crystallography*, 62(Pt 4):439–450, April 2006. ISSN 0907-4449. doi: 10.1107/S0907444906005270.

[86] V. B. Chen, B. W. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography*, 66(1):12–21, January 2010. ISSN 1399-0047. doi: 10.1107/S0907444909042073.

[87] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments

using Clustal Omega. *Molecular Systems Biology*, 7(1):539–539, April 2014. ISSN 1744-4292. doi: 10.1038/msb.2011.75.

[88] C. S. Bond and A. W. Schttelkopf. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta crystallographica. Section D, Biological crystallography*, 65(5):510–512, May 2009. ISSN 1399-0047. doi: 10.1107/S0907444909007835.

[89] W. L. DeLano. PyMOL. http://www.pymol.org., 2002.

[90] W. F. Studier. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification*, 41(1):207–234, May 2005. ISSN 1046-5928.

[91] M. Piotto, V. Saudek, and V. Sklenr. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *Journal of biomolecular NMR*, 2(6):661–665, November 1992. ISSN 0925-2738.

[92] Gale Rhodes. *Crystallography Made Crystal Clear*. Academic Press, 3rd edition, 2006.

[93] G. L. Taylor. Introduction to phasing. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 4):325–338, April 2010. ISSN 1399-0047. doi: 10.1107/S0907444910006694.

[94] H. Walden. Selenium incorporation using recombinant techniques. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):352–357, April 2010. ISSN 0907-4449. doi: 10.1107/S0907444909038207.

[95] E. Krissinel and K. Henrick. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D Biological Crystallography*, 60(12):2256–2268, December 2004. ISSN 0907-4449. doi: 10.1107/S0907444904026460.

[96] M. Wiederstein, M. Gruber, K. Frank, F. Melo, and M. J. Sippl. Structure-based characterization of multiprotein complexes. *Structure (London, England: 1993)*, 22(7):1063–1070, July 2014. ISSN 1878-4186. doi: 10.1016/j.str.2014.05.005.

[97] E. Krissinel and K. Henrick. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, 372(3):774–797, September 2007. ISSN 0022-2836. doi: 10.1016/j.jmb.2007.05.022.

[98] T. Panavas, C. Sanders, and T. R. Butt. SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. *Methods in molecular biology (Clifton, N.J.)*, 497:303–317, 2009. ISSN 1064-3745. doi: 10.1007/978-1-59745-566-4_20.

[99] D. S. Nielsen, H. N. Hoang, R. Lohman, Timothy A. Hill, A. J. Lucke, D. J. Craik, D. J. Edmonds, D. A. Griffith, C. J. Rotter, R. B. Ruggeri, D. A. Price, S. Liras, and D. P. Fairlie. Improving on Nature: Making a Cyclic Heptapeptide Orally Bioavailable. *Angewandte Chemie International Edition*, 53(45):12059–12063, November 2014. ISSN 1521-3773. doi: 10.1002/anie.201405364.

[100] J. A. McIntosh, Z. Lin, M. D. B. Tianero, and E. W. Schmidt. Aestuaramides, a Natural Library of Cyanobactin Cyclic Peptides Resulting from Isoprene-Derived Claisen Rearrangements. *ACS Chemical Biology*, 8(5):877–883, May 2013. ISSN 1554-8929. doi: 10.1021/cb300614c.

[101] K. L. Dunbar, J. R. Chekan, C. L. Cox, B. J. Burkhart, S. K. Nair, and D. A. Mitchell. Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nature Chemical Biology*, 10(10):823–829, October 2014. ISSN 1552-4450. doi: 10.1038/nchembio.1608.

[102] E. Blumenthal and J. B. M. Herbert. Interchange reactions of oxygen. 1. Interchange of oxygen between water and potassium phosphate in solution. *Transactions of the Faraday Society*, 33(0):849–852, January 1937. ISSN 0014-7672. doi: 10.1039/TF9373300849.

[103] G. A. Miller, S. Rosenzweig, and R. L. Switzer. Oxygen-18 studies of the mechanism of pyrophosphoryl group transfer catalyzed by phosphoribosylpyrophosphate synthetase. *Archives of Biochemistry and Biophysics*, 171(2):732–736, December 1975. ISSN 0003-9861.

[104] M. A. Ortega, Y. Hao, Q. Zhang, M. C. Walker, W. A. van der Donk, and S. K. Nair. Structure and mechanism of the tRNA-dependent lantibiotic dehydratase

NisB. *Nature*, 517(7535):509–512, January 2015. ISSN 1476-4687. doi: 10.1038/nature13888.

[105] J. McConathy and M. J. Owens. Stereochemistry in Drug Action. *Primary Care Companion to The Journal of Clinical Psychiatry*, 5(2):70–73, 2003. ISSN 1523-5998.

[106] L. . Nguyen, H. He, and C. Pham-Huy. Chiral Drugs: An Overview. *International Journal of Biomedical Science : IJBS*, 2(2):85–100, June 2006. ISSN 1550-9702.

[107] T. Nakanishi, N. Yamakawa, T. Asahi, N. Shibata, B. Ohtani, and T. Osaka. Chiral discrimination between thalidomide enantiomers using a solid surface with two-dimensional chirality. *Chirality*, 16(S1):S36–S39, January 2004. ISSN 1520-636X. doi: 10.1002/chir.20039.

[108] M. Sattler, J. Schleucher, and C. Griesinger. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 34 (2):93–158, March 1999. ISSN 0079-6565. doi: 10.1016/S0079-6565(98)00025-9.

[109] M. Jaspars. The origins of cyanobactin chemistry and biology. *Chemical Communications*, 50(71):10174–10176, August 2014. ISSN 1364-548X. doi: 10.1039/C3CC49252D.

[110] S. V. Ambudkar, S. Dey, C. A. Hrycyna, M. Ramachandra, I. Pastan, and M. M. Gottesman. Biochemical, cellular, and pharmacological aspects of the multidrug transporter. *Annual Review of Pharmacology and Toxicology*, 39:361–398, 1999. ISSN 0362-1642. doi: 10.1146/annurev.pharmtox.39.1.361.

[111] C. Pauli-Magnus, O. von Richter, O. Burk, A. Ziegler, T. Mettang, M. Eichelbaum, and M. F. Fromm. Characterization of the major metabolites of verapamil as substrates and inhibitors of P-glycoprotein. *The Journal of Pharmacology and Experimental Therapeutics*, 293(2):376–382, May 2000. ISSN 0022-3565.

[112] M. S. Donia and E. W Schmidt. Linking chemistry and genetics in the growing cyanobactin natural products family. *Chemistry & biology*, 18(4):508–519, April 2011. ISSN 1879-1301. doi: 10.1016/j.chembiol.2011.01.019.

[113] C. K. Larive, D. Jayawickrama, and L. Orfi. Quantitative Analysis of Peptides with NMR Spectroscopy. *Applied Spectroscopy*, 51(10):1531–1536, October 1997. ISSN 0003-7028, 1943-3530. doi: 10.1366/0003702971939055.

[114] U. Holzgrabe. Quantitative NMR spectroscopy in pharmaceutical applications. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 57(2):229–240, August 2010. ISSN 1873-3301. doi: 10.1016/j.pnmrs.2010.05.001.

[115] T. Rundlf, M. Mathiasson, S. Bekiroglu, B. Hakkarainen, T. Bowden, and T. Arvidsson. Survey and qualification of internal standards for quantification by 1h NMR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis*, 52 (5):645–651, September 2010. ISSN 1873-264X. doi: 10.1016/j.jpba.2010.02.007.

[116] S. Akoka, L. Barantin, and M. Trierweiler. Concentration Measurement by Proton NMR Using the ERETIC Method. *Analytical Chemistry*, 71(13):2554–2557, July 1999. ISSN 0003-2700. doi: 10.1021/ac981422i.

[117] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch. Protein Identification and Analysis Tools on the ExPASy Server. In JohnM. Walker, editor, *The Proteomics Protocols Handbook*, pages 571–607. Humana Press, January 2005. ISBN 978-1-58829-343-5. DOI: 10.1385/1-59259-890-0:571.

[118] R. W. Robey, S. Shukla, E. M. Finley, R. K. Oldham, D. Barnett, S. V. Ambudkar, T. Fojo, and S. E. Bates. Inhibition of P-glycoprotein (ABCB1)- and multidrug resistance-associated protein 1 (ABCC1)-mediated transport by the orally administered inhibitor, CBT-1((R)). *Biochemical Pharmacology*, 75(6): 1302–1312, March 2008. ISSN 1873-2968. doi: 10.1016/j.bcp.2007.12.001.

[119] F. Tiberghien, T. Wenandy, and F. Loor. The potent immunosuppressive cyclosporin FR901459 inhibits the human P-glycoprotein and formyl peptide receptor functions. *The Journal of Antibiotics*, 53(5):509–515, May 2000. ISSN 0021-8820.

[120] A. B. Ke, S. Eyal, F. S. Chung, J. M. Link, D. A. Mankoff, M. Muzi, and J. D. Unadkat. Modeling cyclosporine A inhibition of the distribution of a P-glycoprotein PET ligand, 11c-verapamil, into the maternal brain and fetal liver of the pregnant nonhuman primate: impact of tissue blood flow and site of inhibition.

*Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine*, 54(3):437–446, March 2013. ISSN 1535-5667. doi: 10.2967/jnumed.112.111732.

# Appendix A - Media and Buffer Compositions

| | |
|---|---|
| Minimal medium | 19 mM $(NH_4)_2SO_4$, 22 mM $KH_2PO_4$, 22 mM $Na_2HPO_4$, 9 mM NaCl |
| Autoinduction medium | ZY (10 g N-Z-amine AS, 5 g yeast extract, 925 mL $H_2O$ / L culture). Media supplemented with 1 mM $MgSO_4$, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, 25 mM $(NH_4)_2SO_4$, 50 mM $KH_2PO_4$, 50 mM $Na_2HPO_4$ |
| G-DUF lysis buffer | 150 mM NaCl, 20 mM Bis-tris pH 6.8, 20 mM imidazole pH 8.0, 0.1 % Triton X-100, 3 mM BME |
| G-DUF elution buffer | 150 mM NaCl, 20 mM Bis-tris pH 6.8, 250 mM imidazole pH 8.0, 0.1 % Triton X-100, 3 mM BME |
| Desalt buffer | 150 mM NaCl, 20 mM Bis-tris pH 6.8, 20 mM imidazole pH 8.0, 3 mM BME |
| Desalt buffer 2 | 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME |
| Gel filtration buffer | 150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP |
| Urea lysis buffer | 8 M Urea, 500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME |

| | |
|---|---|
| Urea elution buffer | 8 M Urea, 500 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0, 3 mM BME |
| General lysis buffer | 500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME |
| General elution buffer | 500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME |
| One-pot storage buffer | 150 mM NaCl, 10 mM Bicine pH 8.5, 1 mM TCEP |
| Peptide column buffer | 150 mM NaCl, 20 mM Bicine pH 8.1 |
| One-pot reaction buffer | 150 mM NaCl, 50 mM Bicine pH 9.0 |

# Appendix B - SAR Cyclic Peptide Characterisation

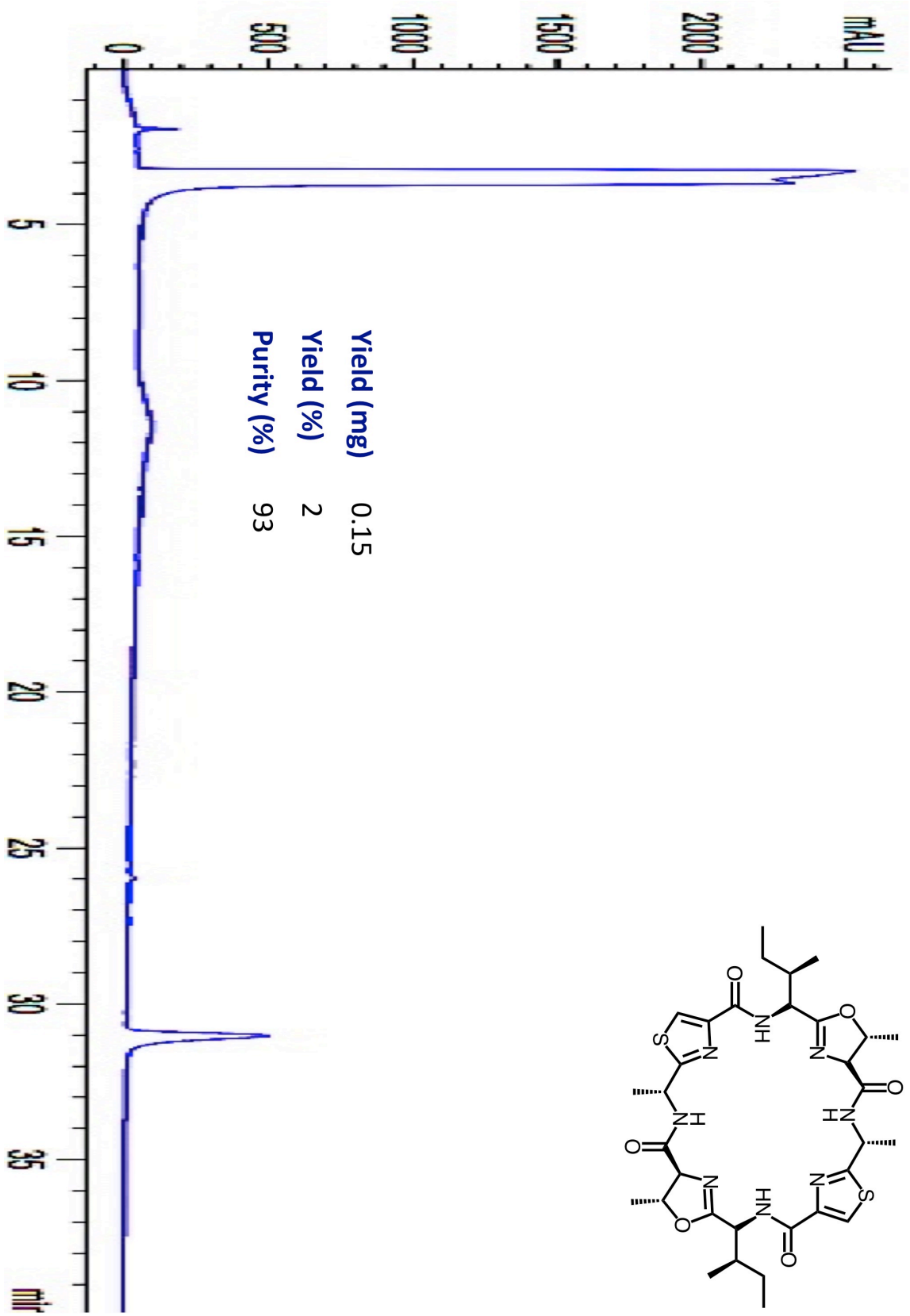





FIGURE 25: **HPLC of cyclo[IT$^{MeOx}$AC$^{Thz}$IT$^{MeOx}$AC$^{Thz}$] (1)**

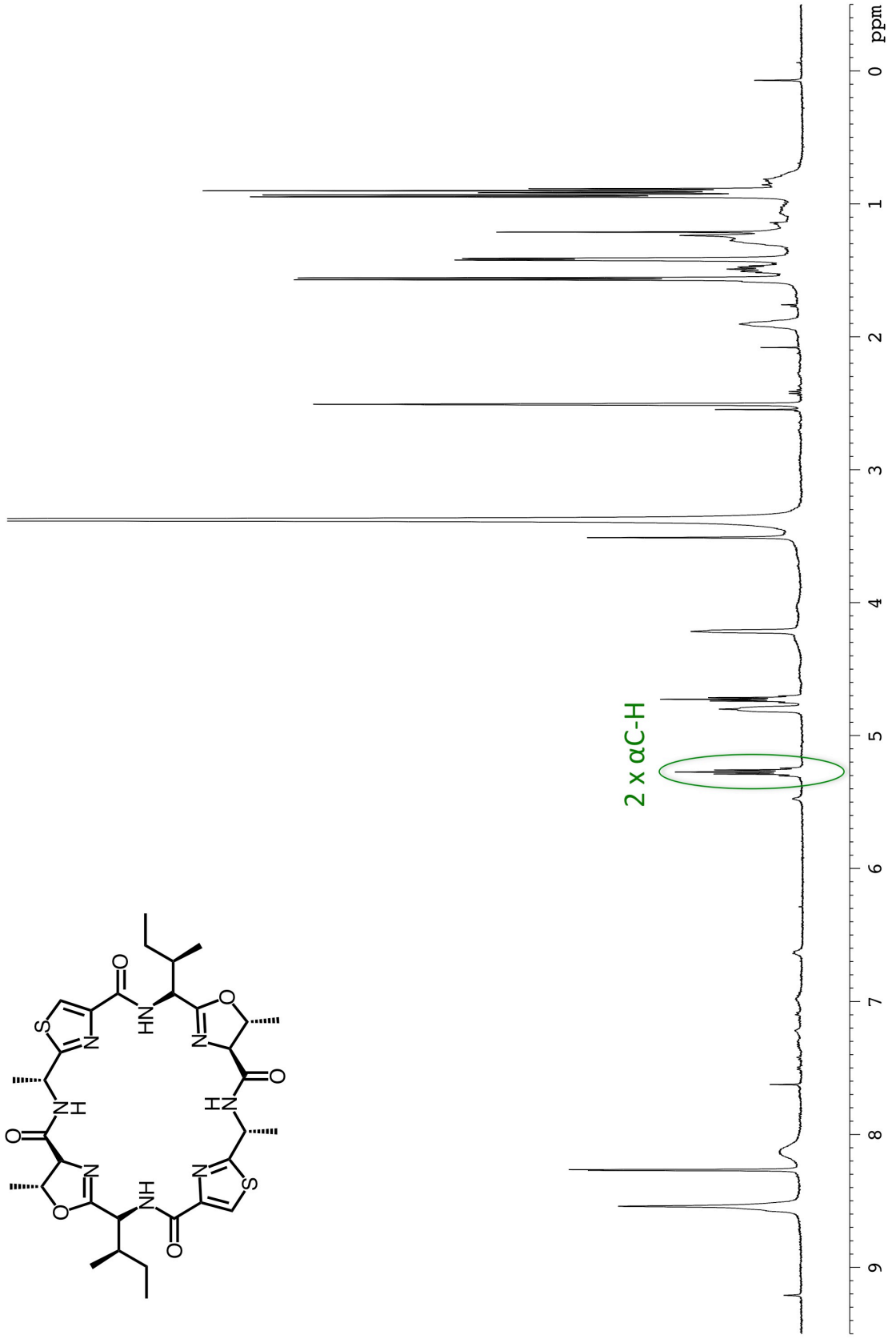



FIGURE 26: **NMR of cyclo[$\mathbf{IT^{MeOx}AC^{Thz}IT^{MeOx}AC^{Thz}}$] (1)**

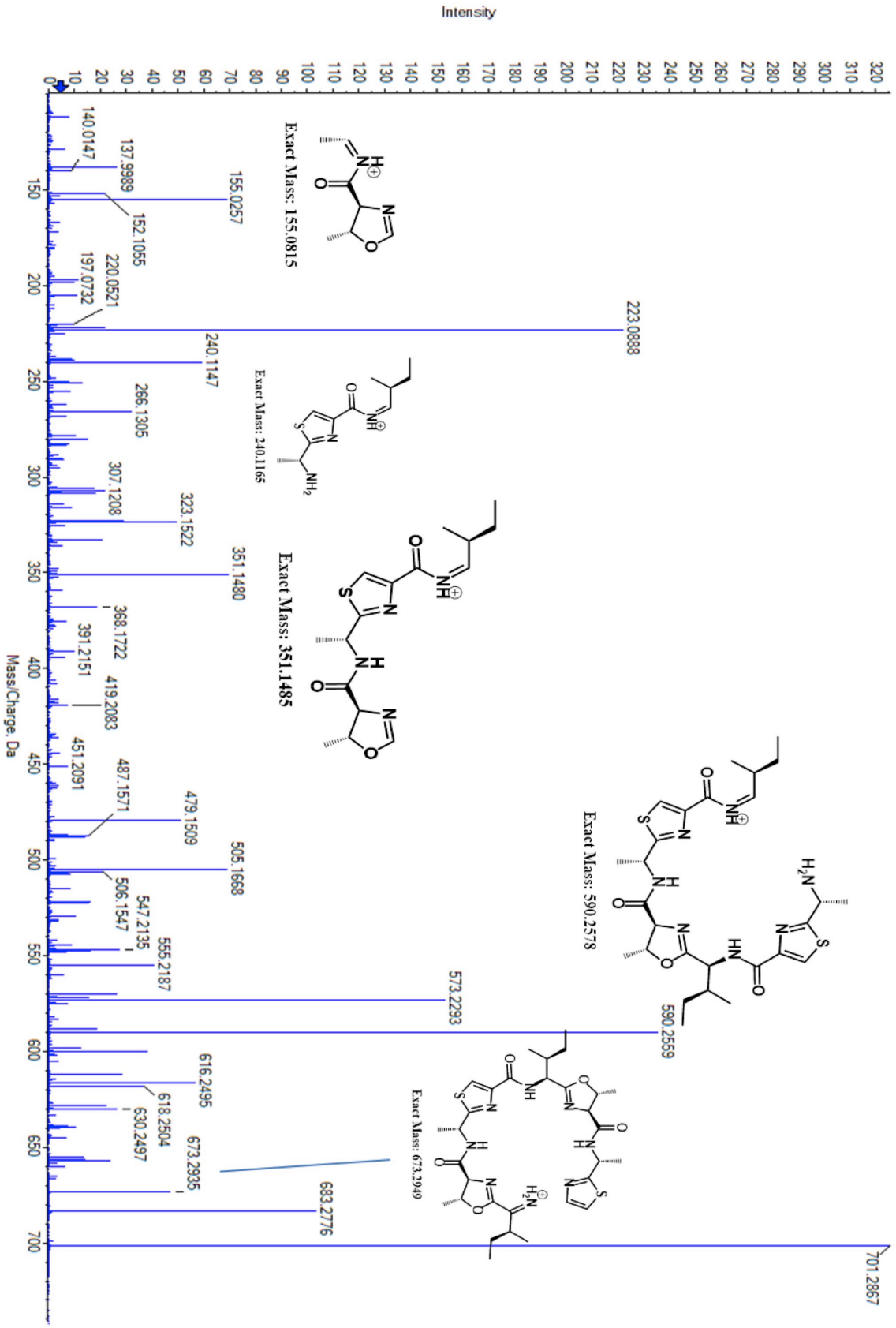



FIGURE 27: **MSMS of cyclo[IT$^{MeOx}$AC$^{Thz}$IT$^{MeOx}$AC$^{Thz}$] (1)**

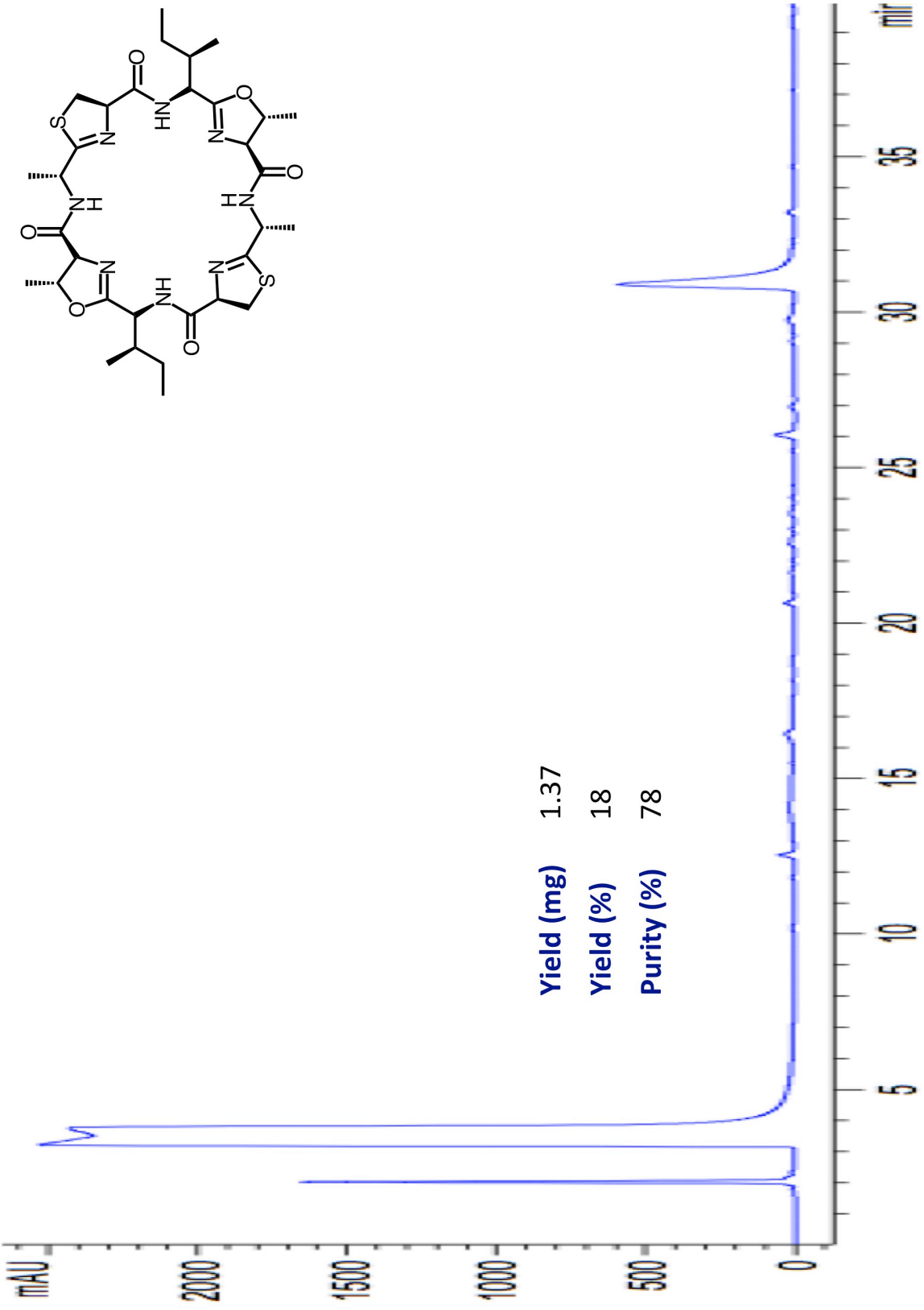



FIGURE 28: **HPLC of cyclo[$\mathbf{IT^{MeOx}AC^{ThH}IT^{MeOx}AC^{ThH}}$] (4)**

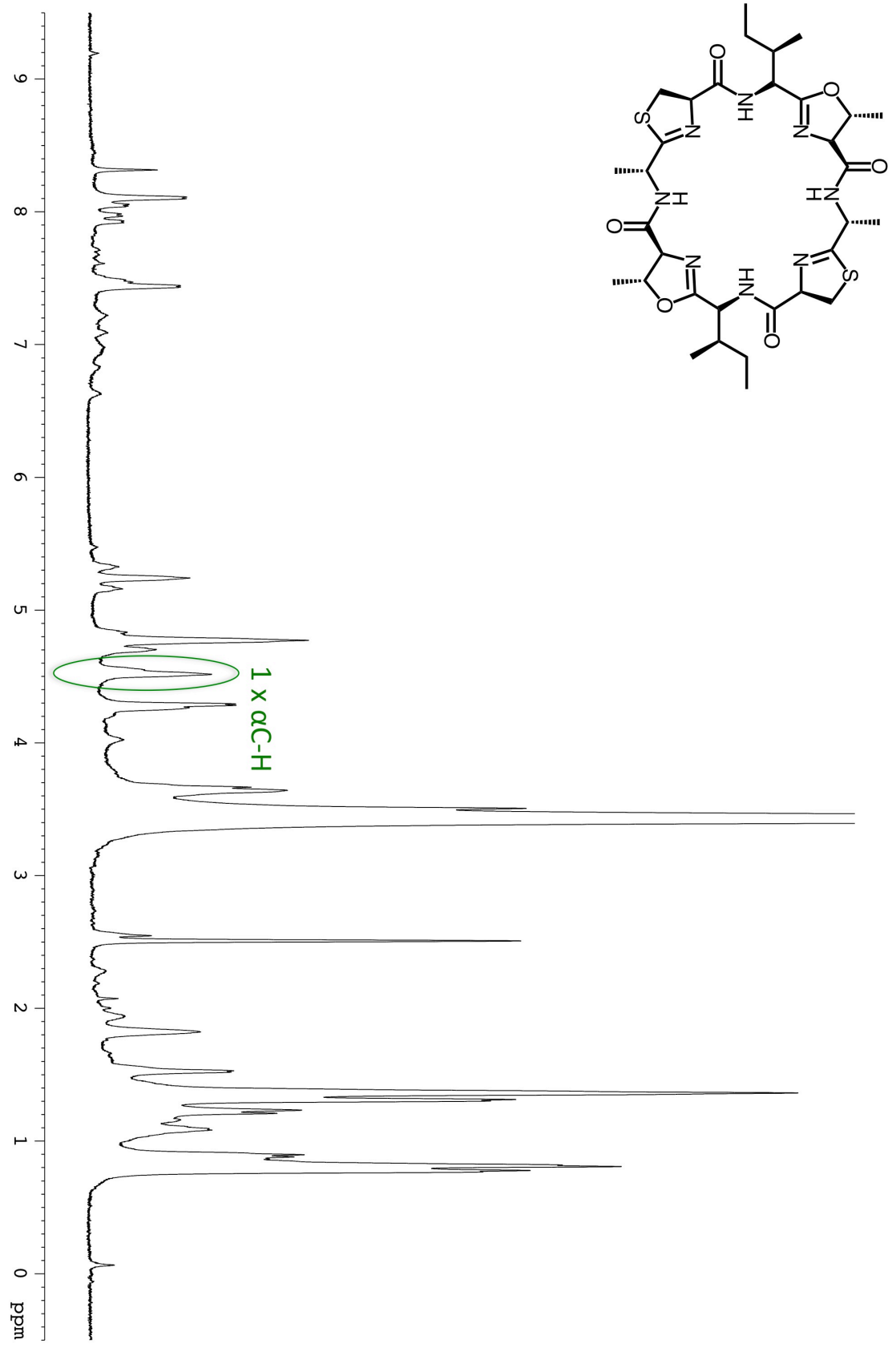



FIGURE 29: **NMR of cyclo[IT$^{MeOx}$AC$^{ThH}$IT$^{MeOx}$AC$^{ThH}$] (4)**

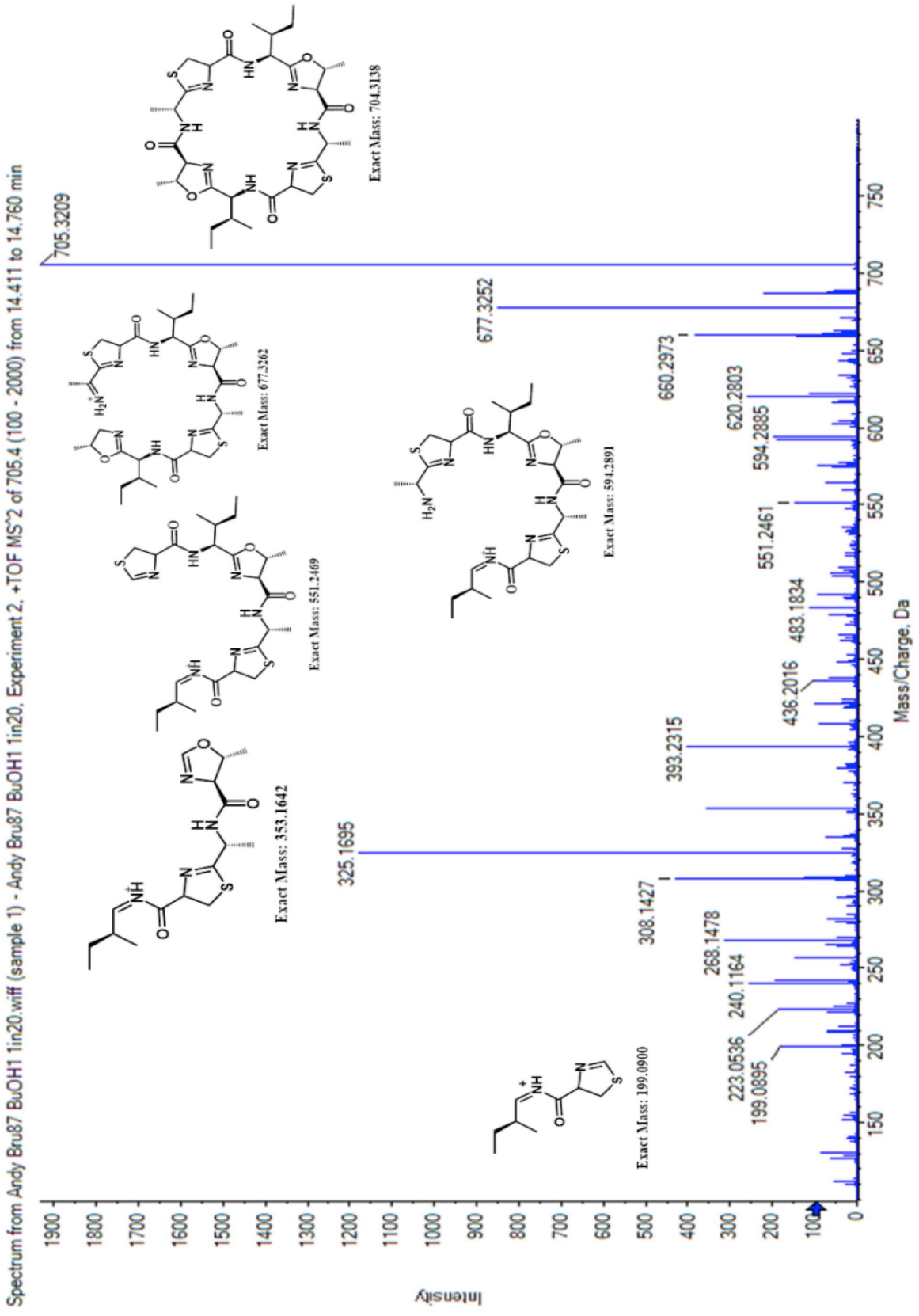



FIGURE 30: **MSMS of cyclo[IT$^{MeOx}$AC$^{ThH}$IT$^{MeOx}$AC$^{ThH}$] (4)**

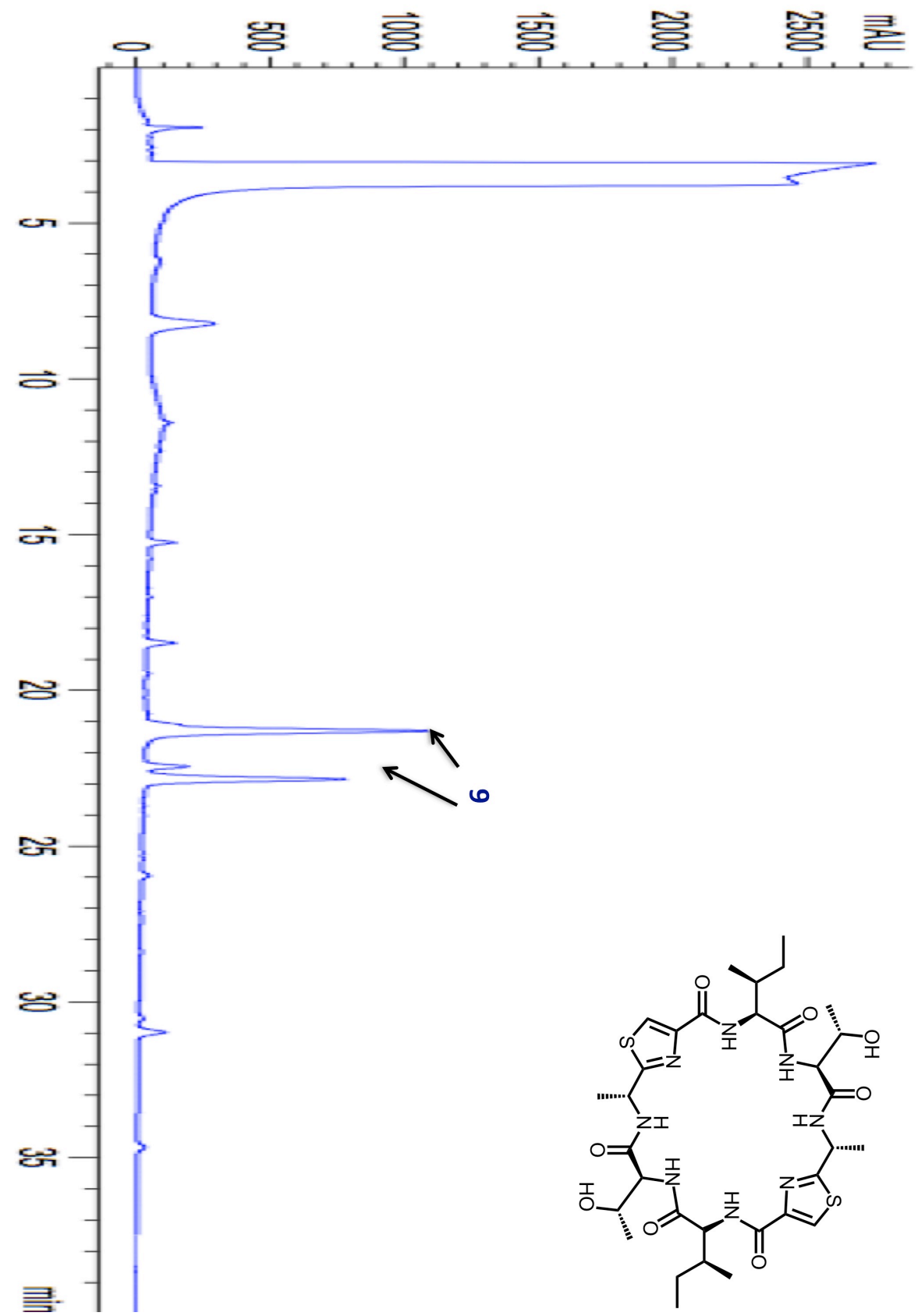



FIGURE 31: **HPLC of cyclo[ITAC$^{Thz}$ITAC$^{Thz}$] (9)**

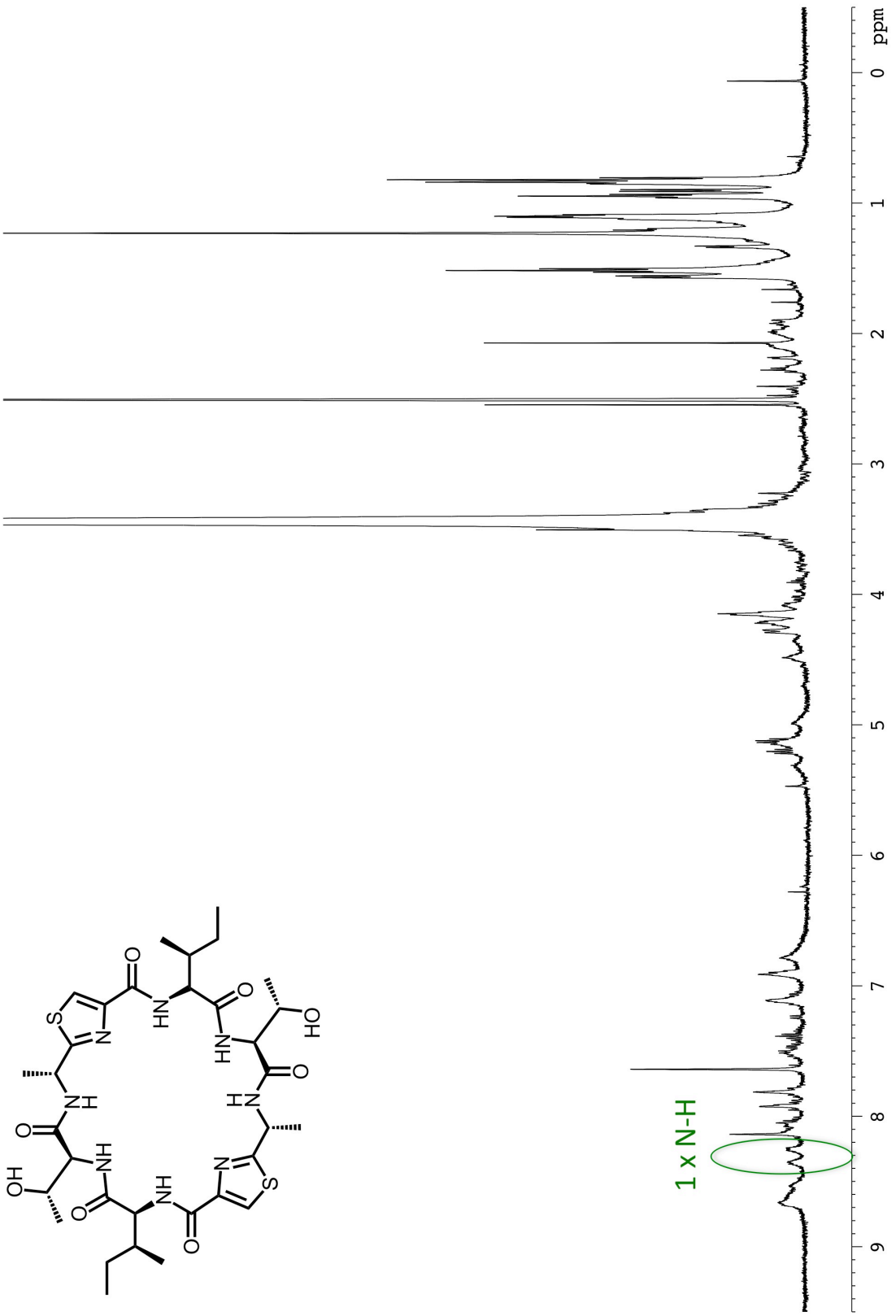



FIGURE 32: **NMR of cyclo[ITAC$^{Thz}$ITAC$^{Thz}$] (9)**

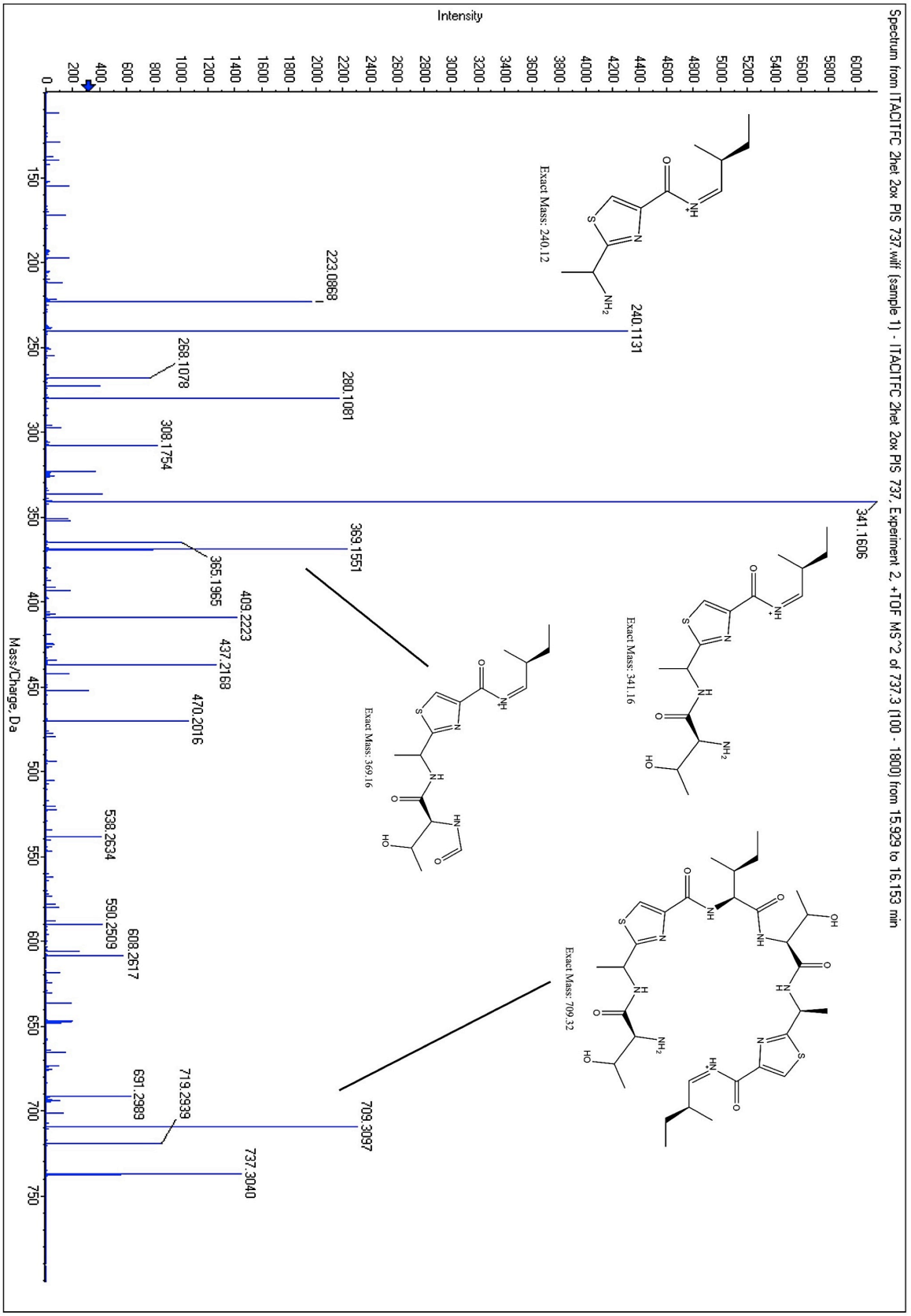



FIGURE 33: **MSMS of cyclo[ITAC$^{Thz}$ITAC$^{Thz}$] (9)**

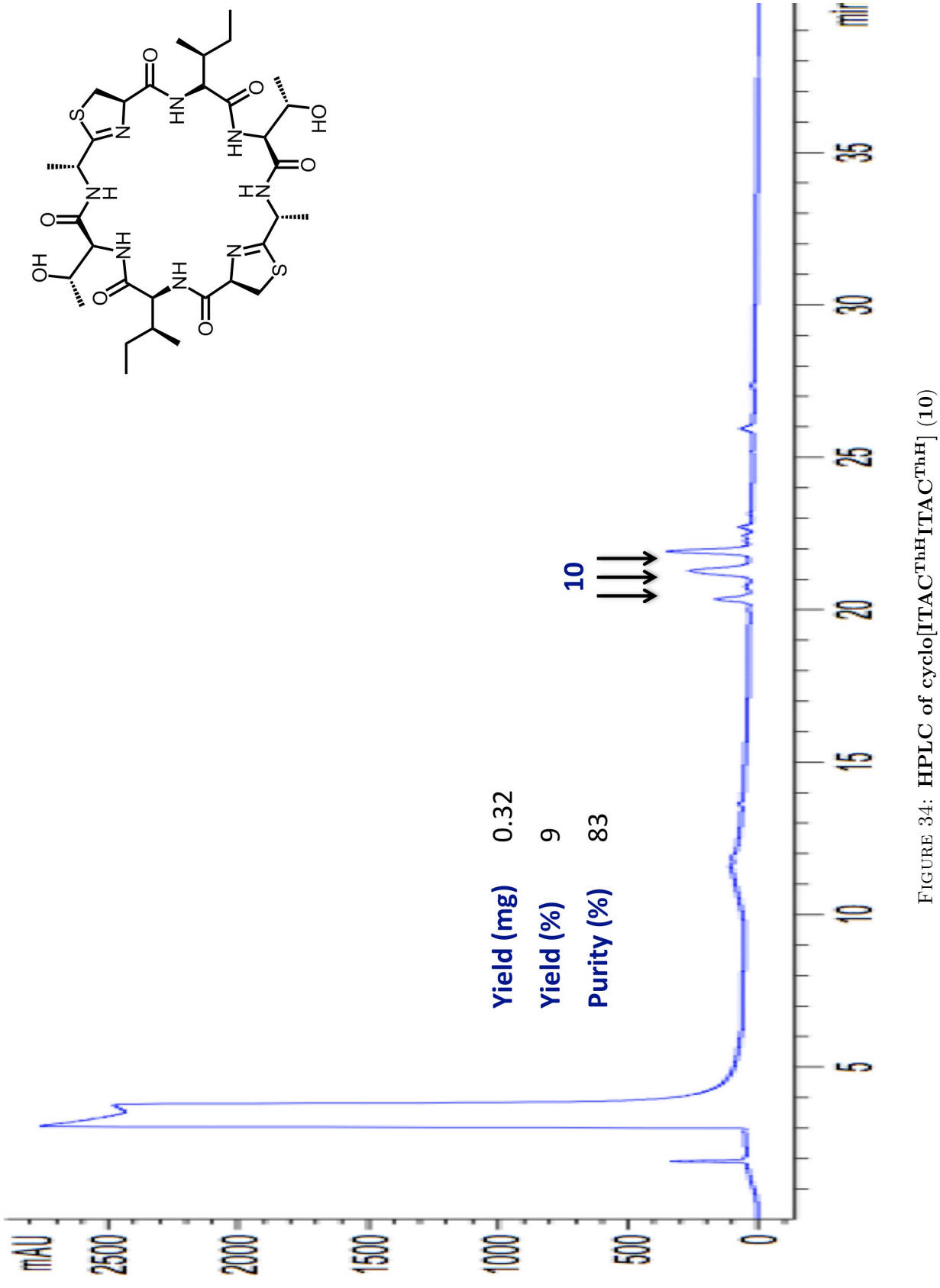



FIGURE 34: **HPLC of cyclo[ITAC$^{ThH}$ITAC$^{ThH}$] (10)**

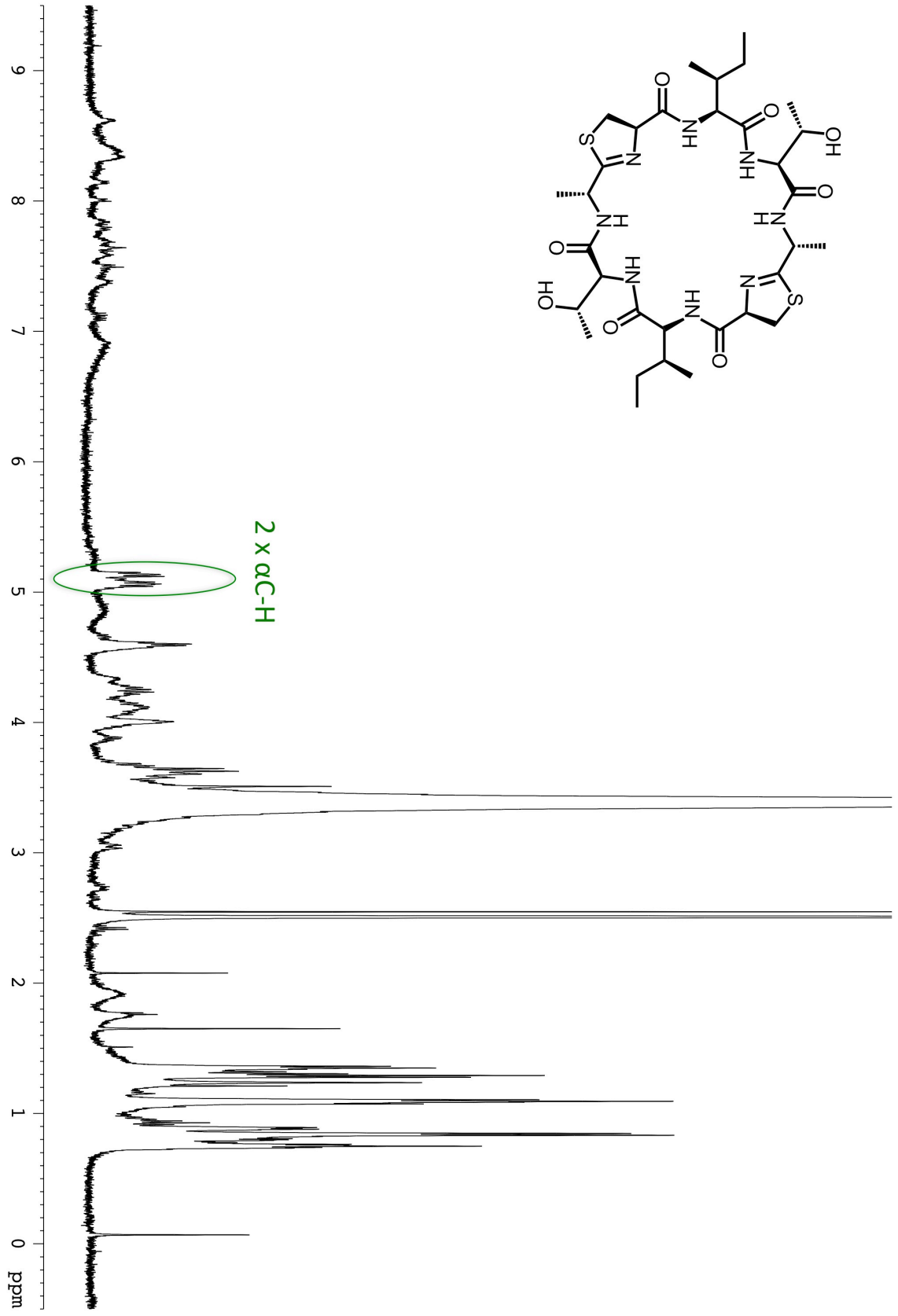



FIGURE 35: **NMR of cyclo[ITAC$^{\text{TbH}}$ITAC$^{\text{TbH}}$] (10)**

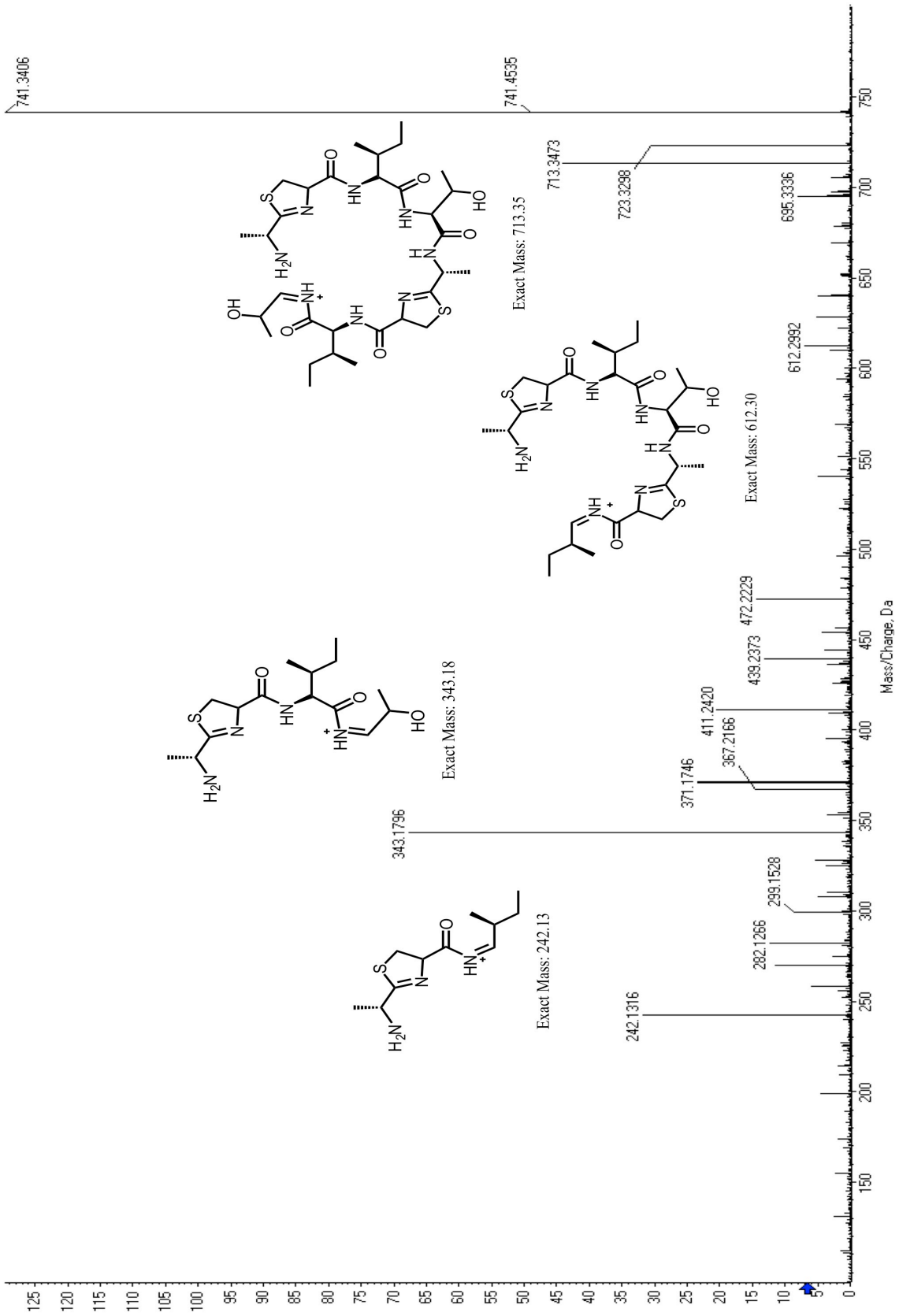



FIGURE 36: **MSMS of cyclo[ITAC$^{\mathbf{ThH}}$ITAC$^{\mathbf{ThH}}$] (10)**

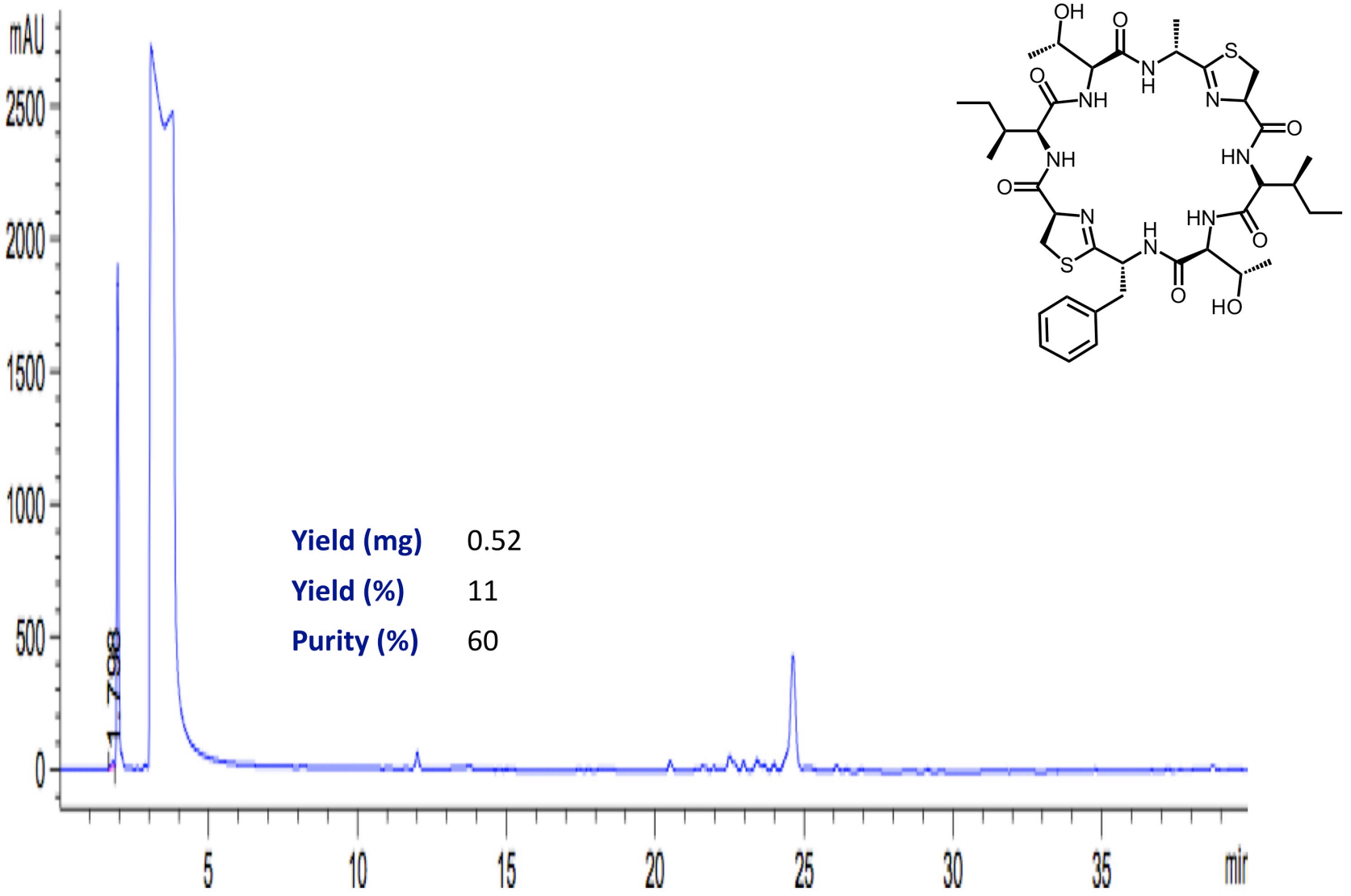



FIGURE 37: **HPLC of cyclo[ITAC$^{ThH}$ITFC$^{ThH}$] (15)**

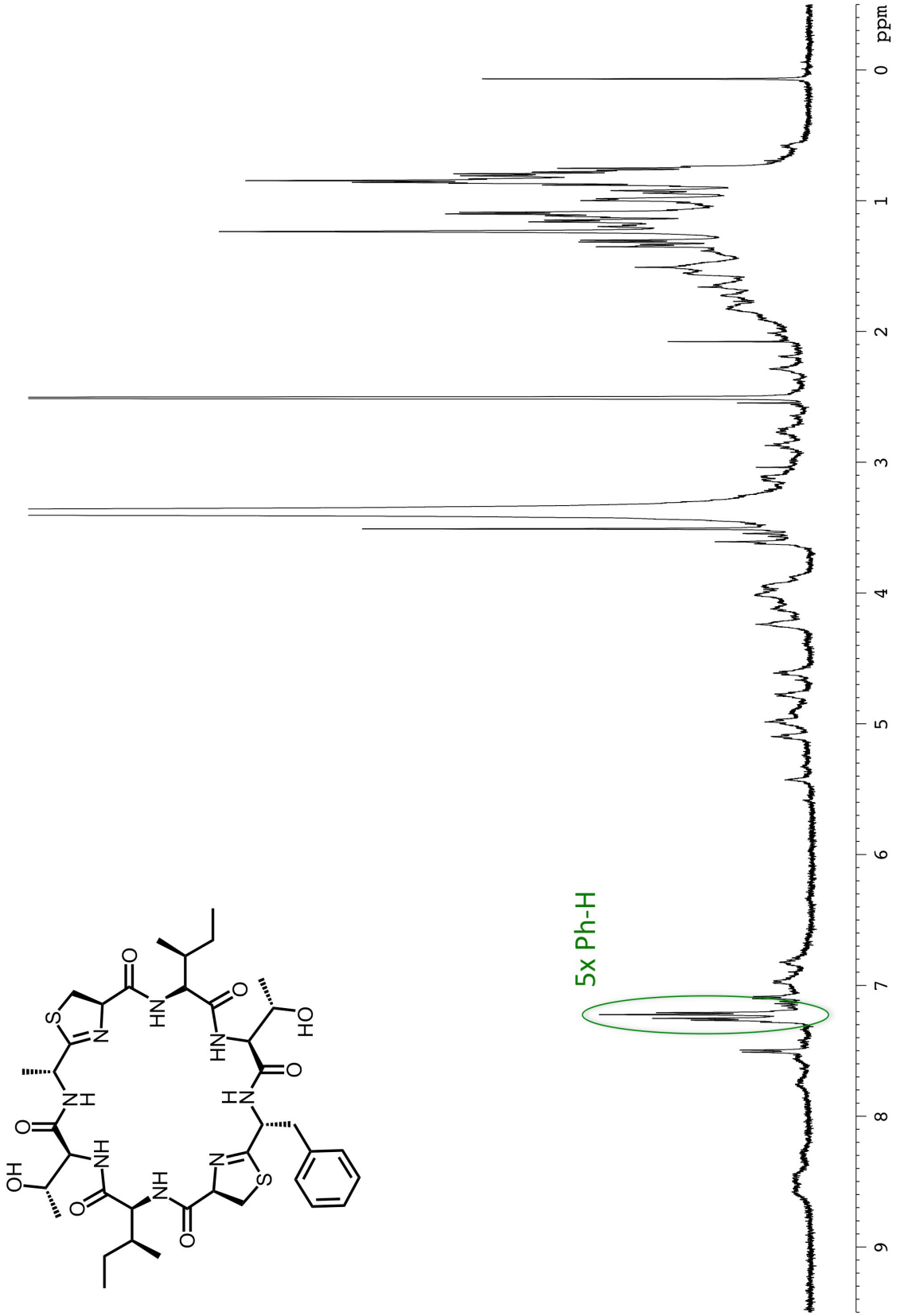



FIGURE 38: **NMR of cyclo[ITAC$^{ThH}$ITFC$^{ThH}$] (15)**

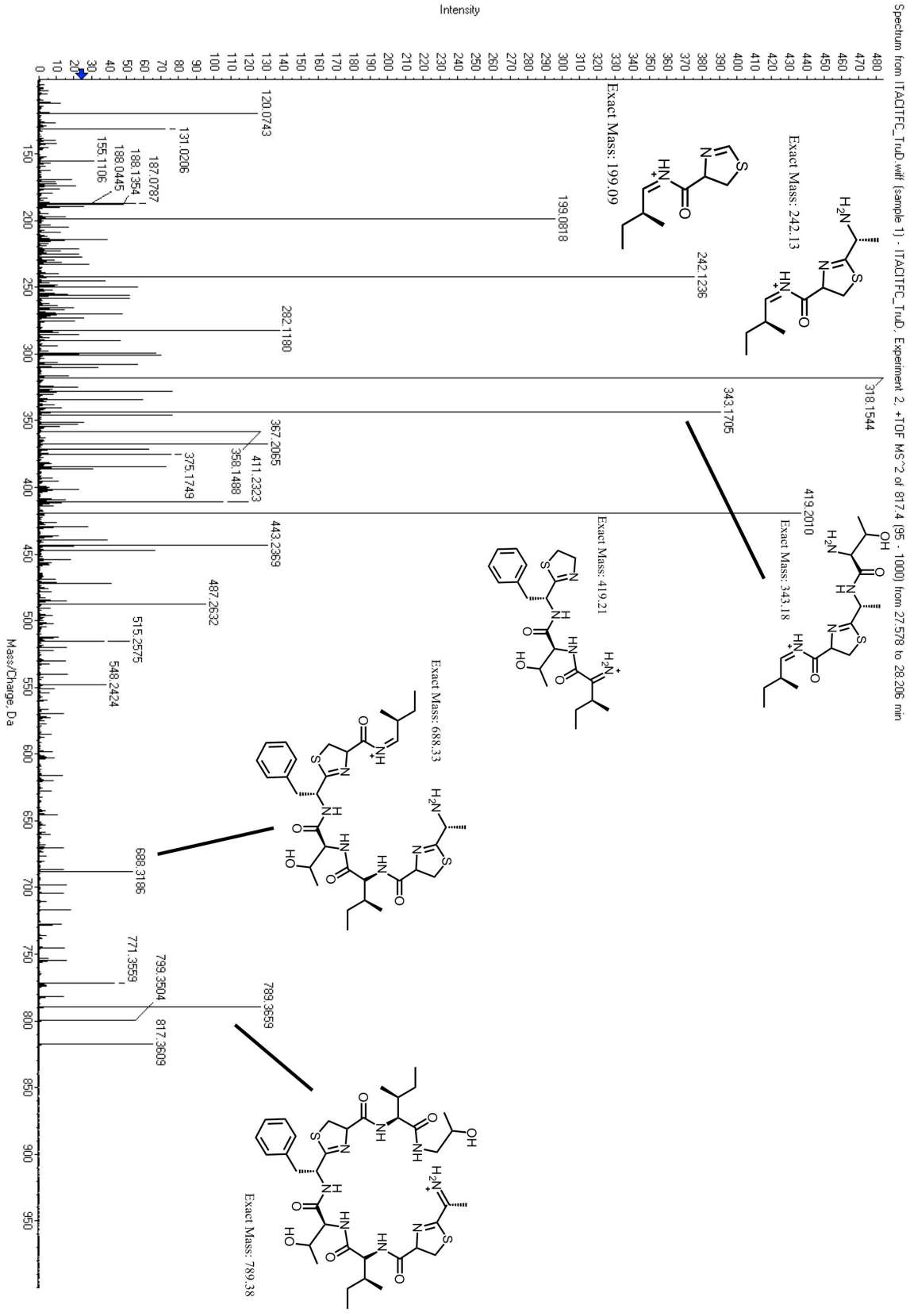



FIGURE 39: **MSMS of cyclo[ITAC$^{ThH}$ITFC$^{ThH}$] (15)**

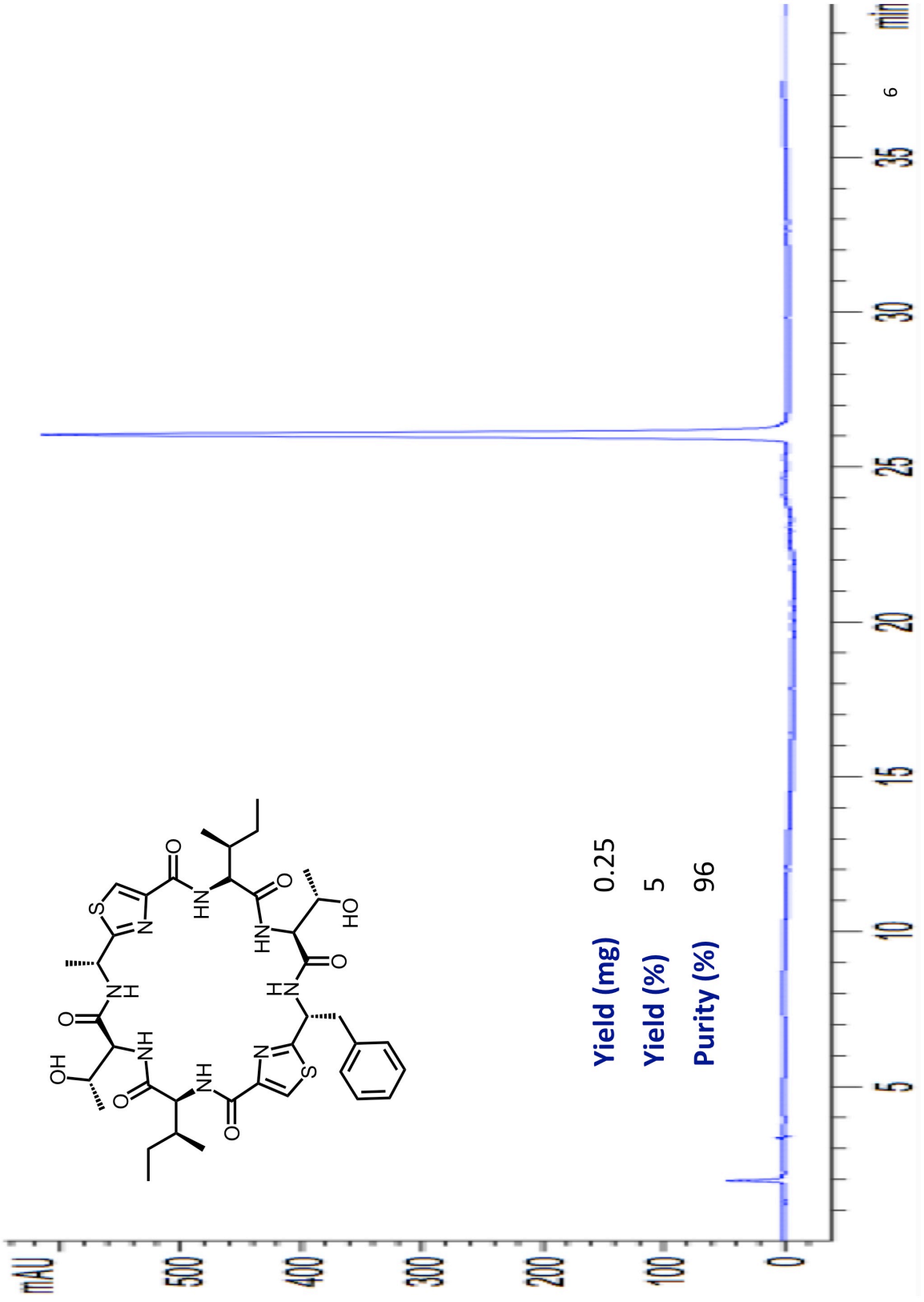



FIGURE 40: **HPLC of cyclo[ITAC$^{Thz}$ITFC$^{Thz}$] (16)**

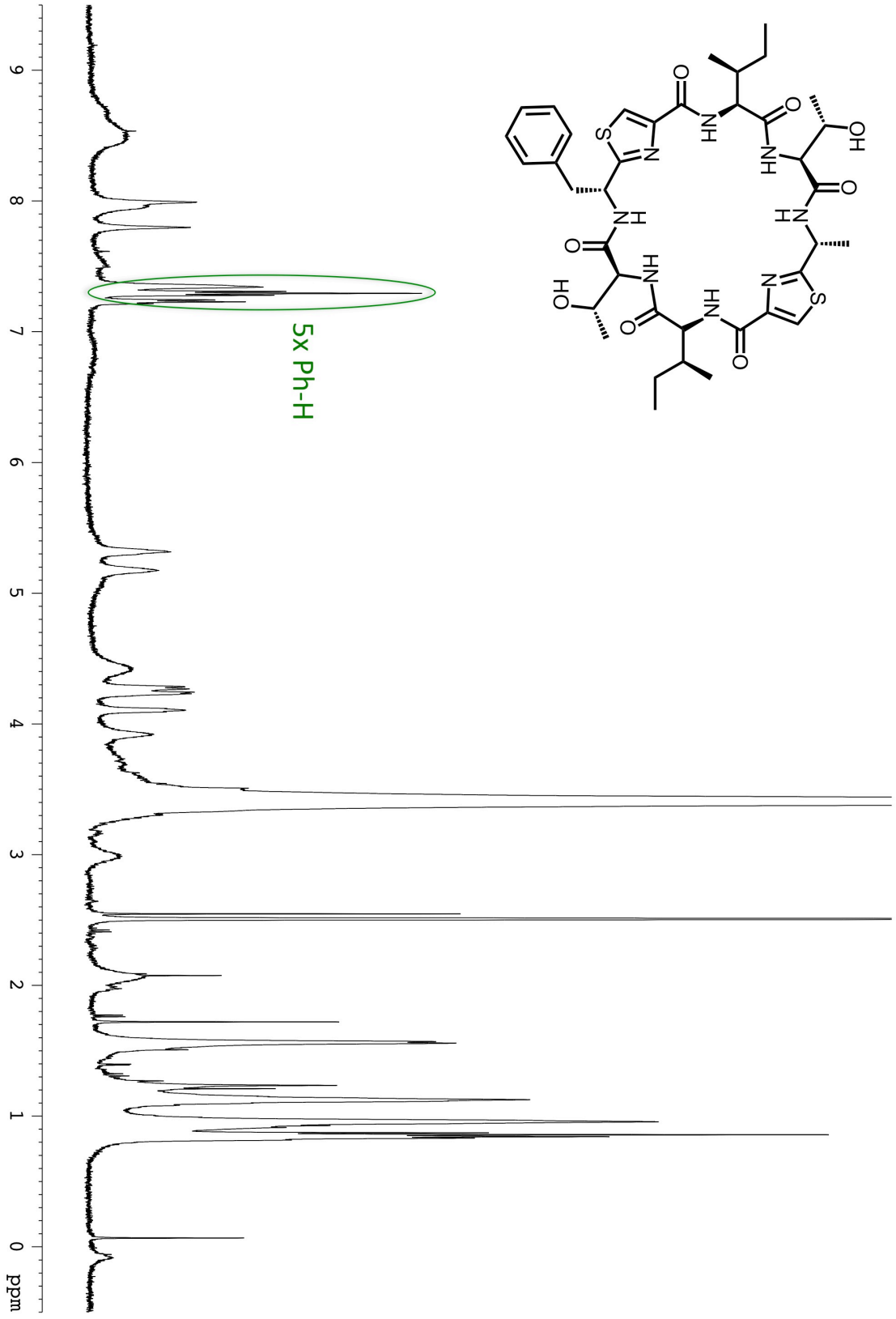



FIGURE 41: **NMR of cyclo[ITAC$^{Thz}$ITFC$^{Thz}$] (16)**

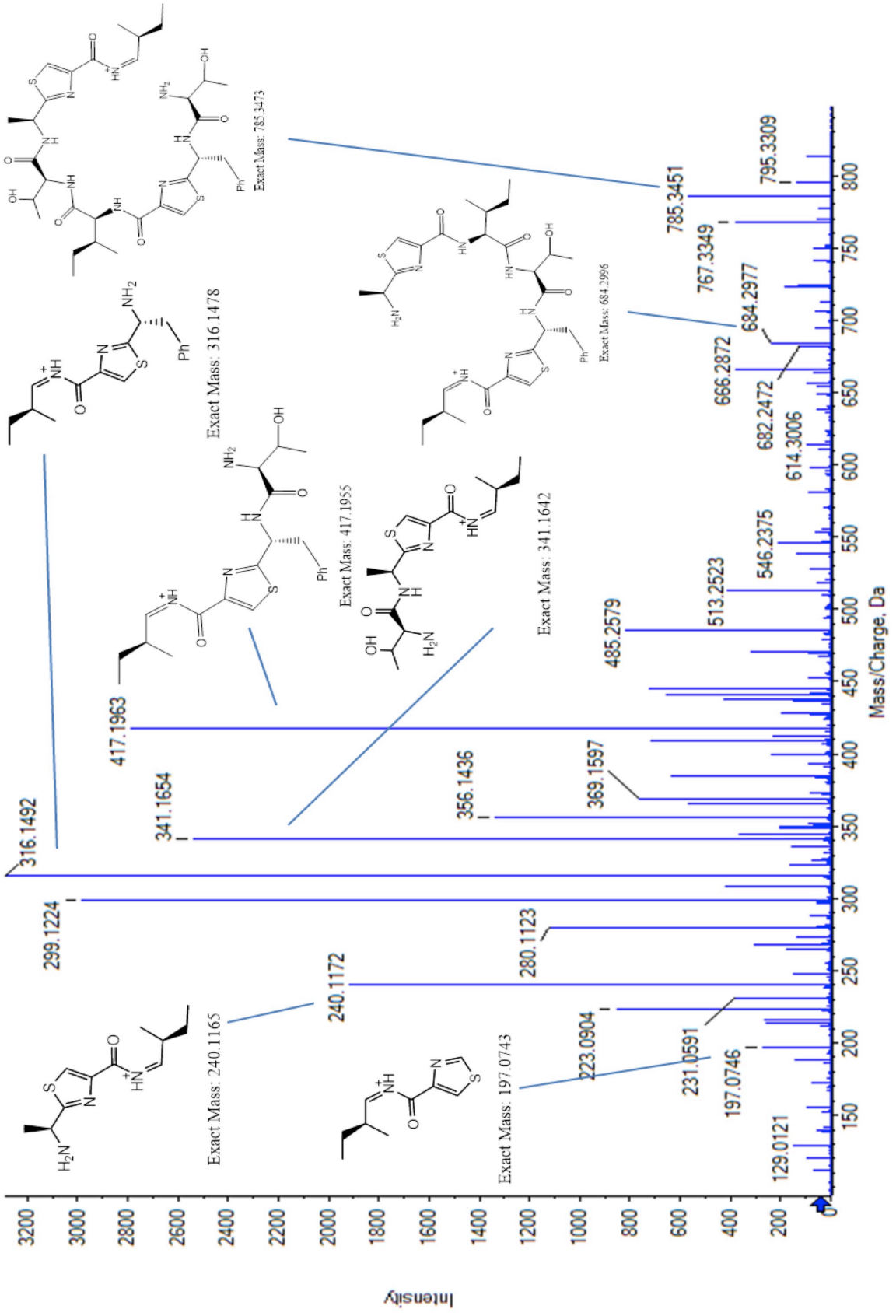



FIGURE 42: **MSMS of cyclo[ITAC$^{Thz}$ITFC$^{Thz}$] (16)**

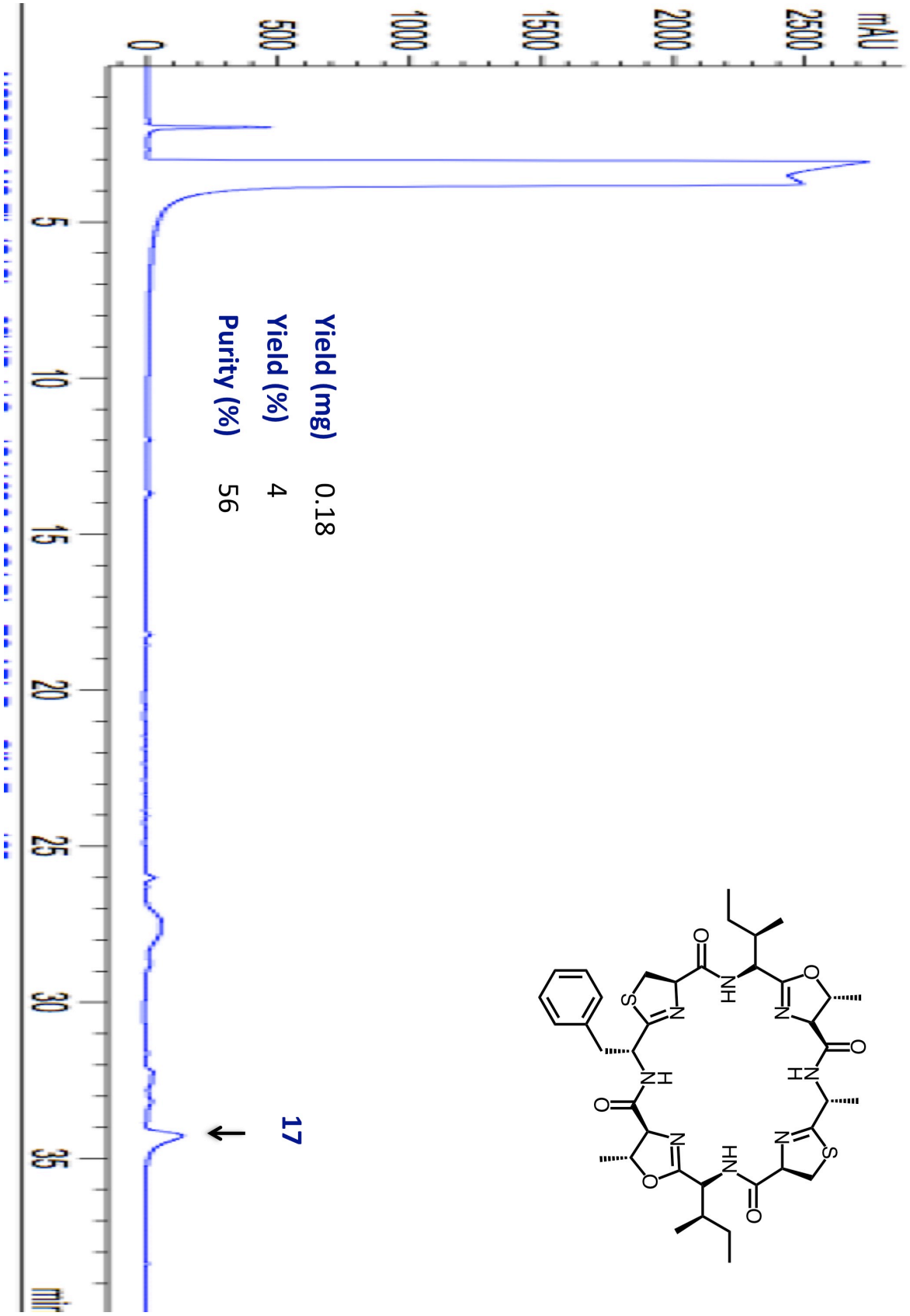



FIGURE 43: **HPLC of cyclo[IT$^{MeOx}$AC$^{ThH}$IT$^{MeOx}$FC$^{ThH}$] (17)**

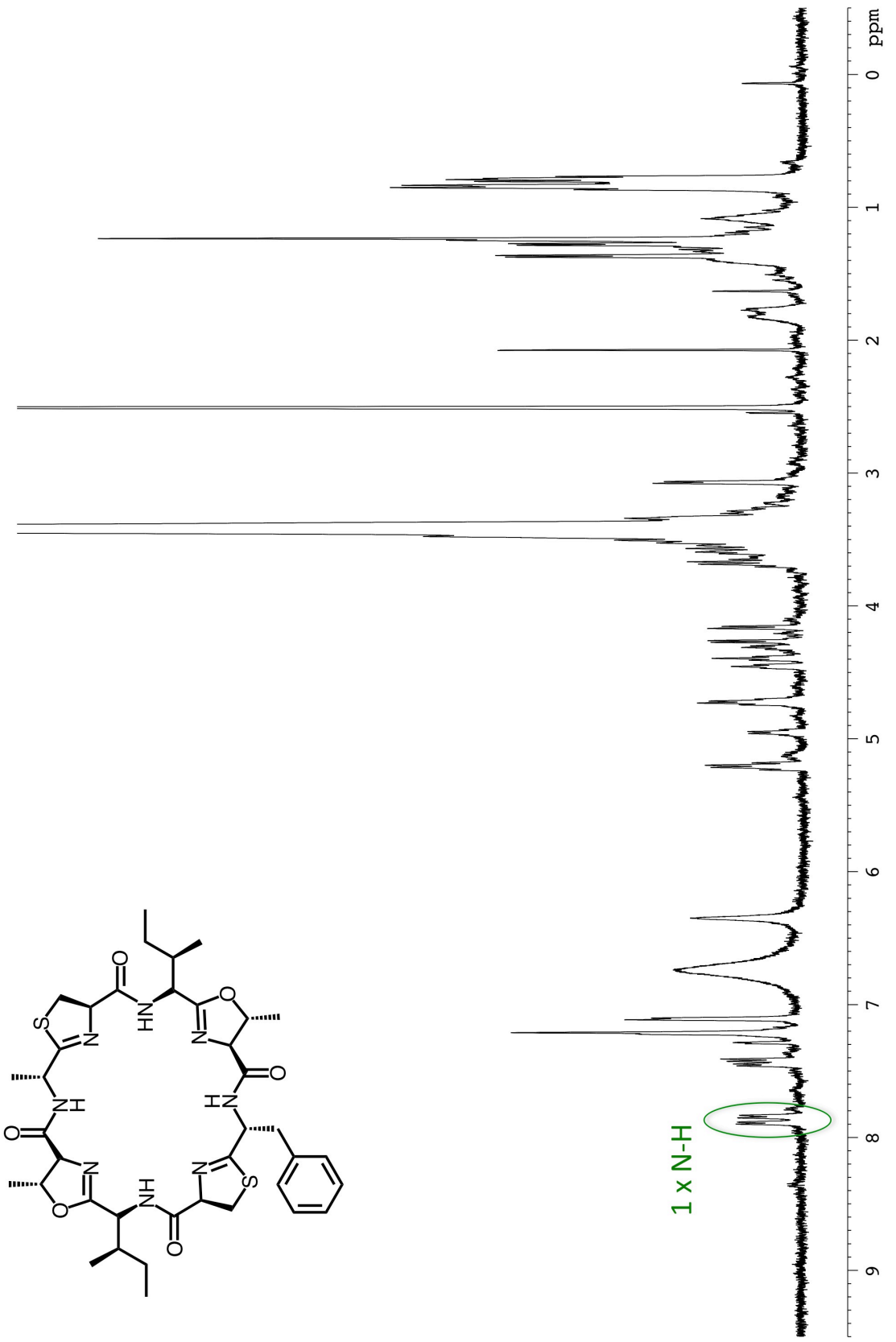



FIGURE 44: **NMR of cyclo[$\mathbf{IT^{MeOx}AC^{ThH}IT^{MeOx}FC^{ThH}}$] (17)**

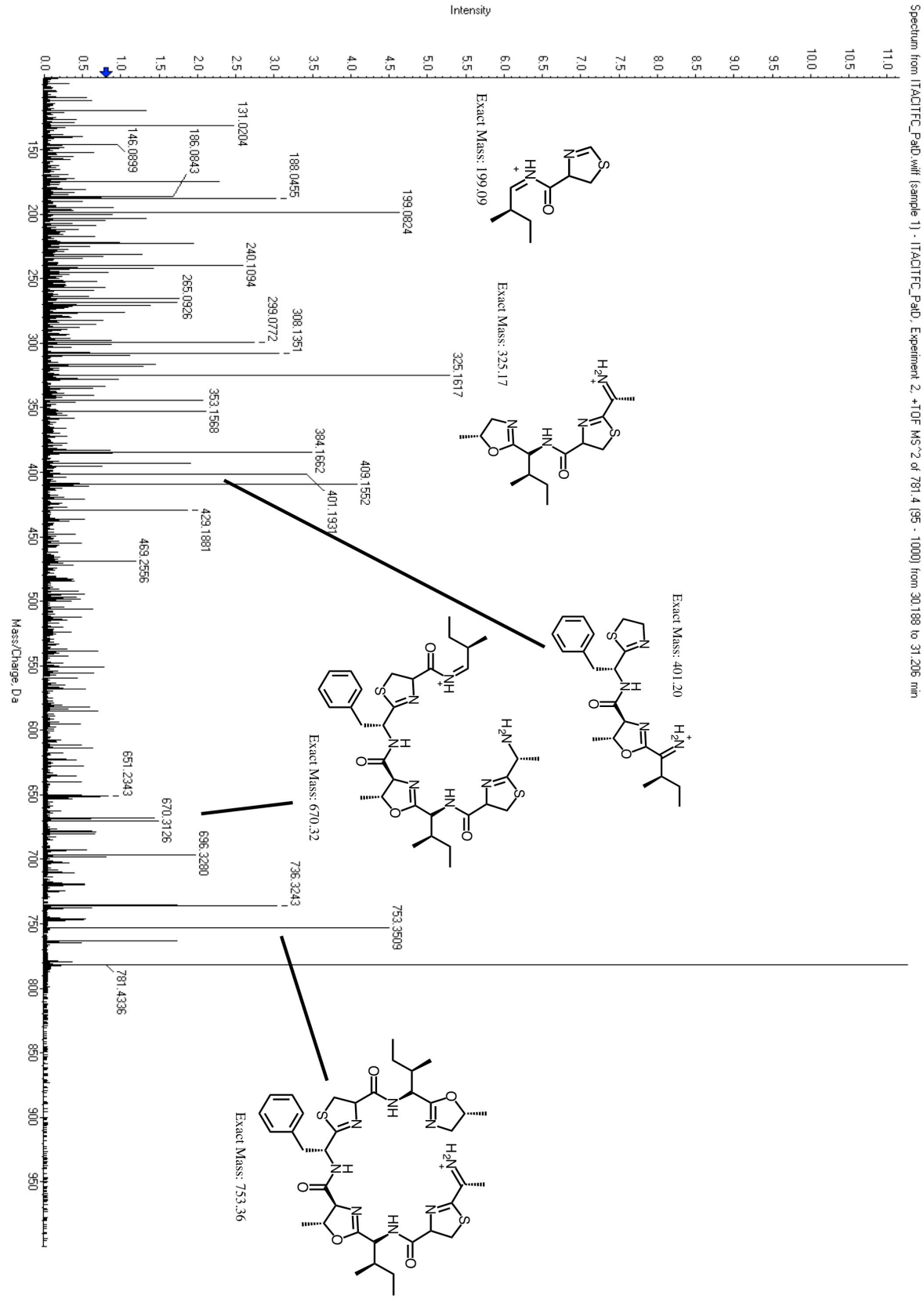



FIGURE 45: **MSMS of cyclo[IT$^{MeOx}$AC$^{ThH}$IT$^{MeOx}$FC$^{ThH}$] (17)**

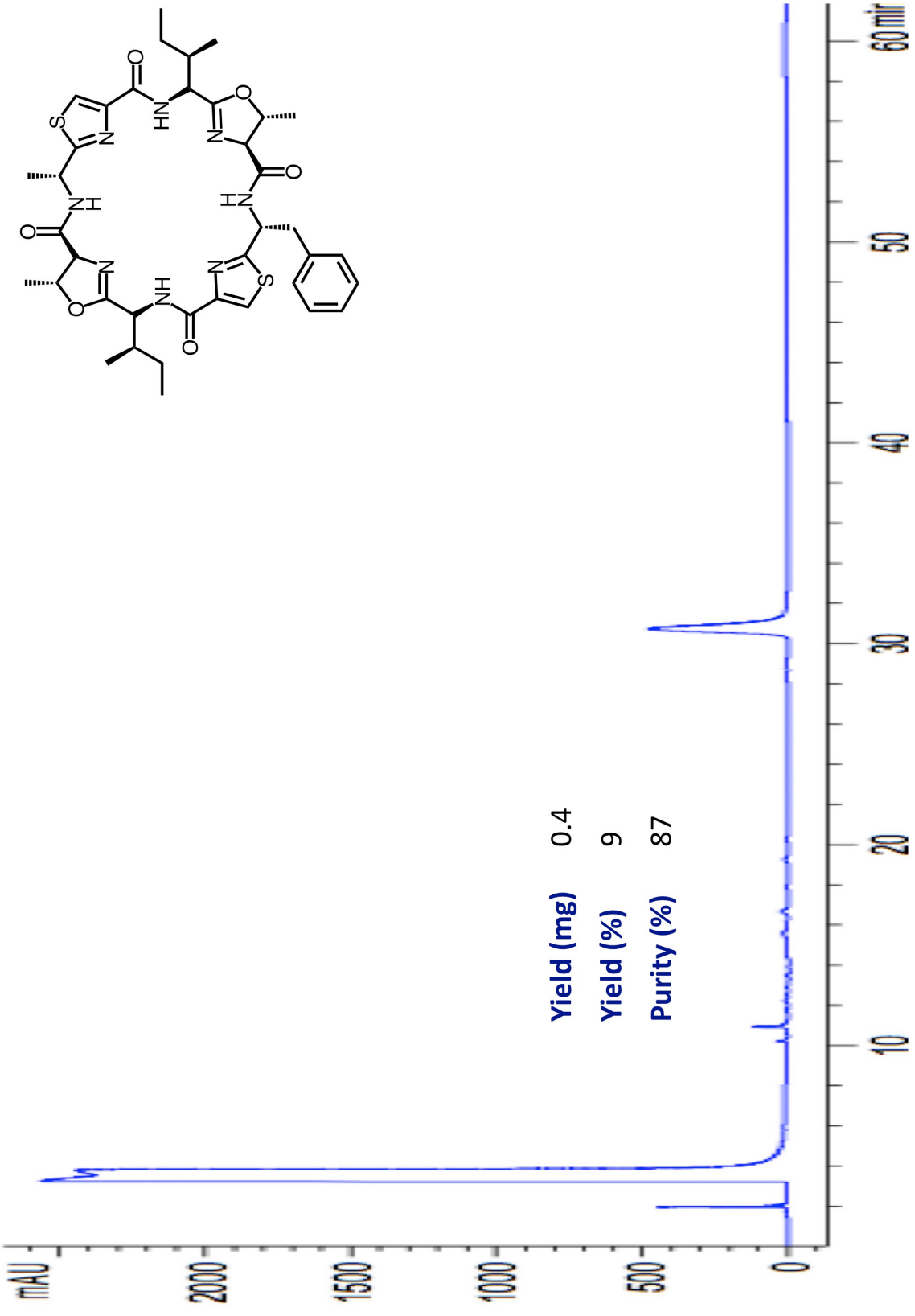



FIGURE 46: **HPLC of cyclo[IT$^{MeOx}$AC$^{Thz}$IT$^{MeOx}$FC$^{Thz}$] (18)**

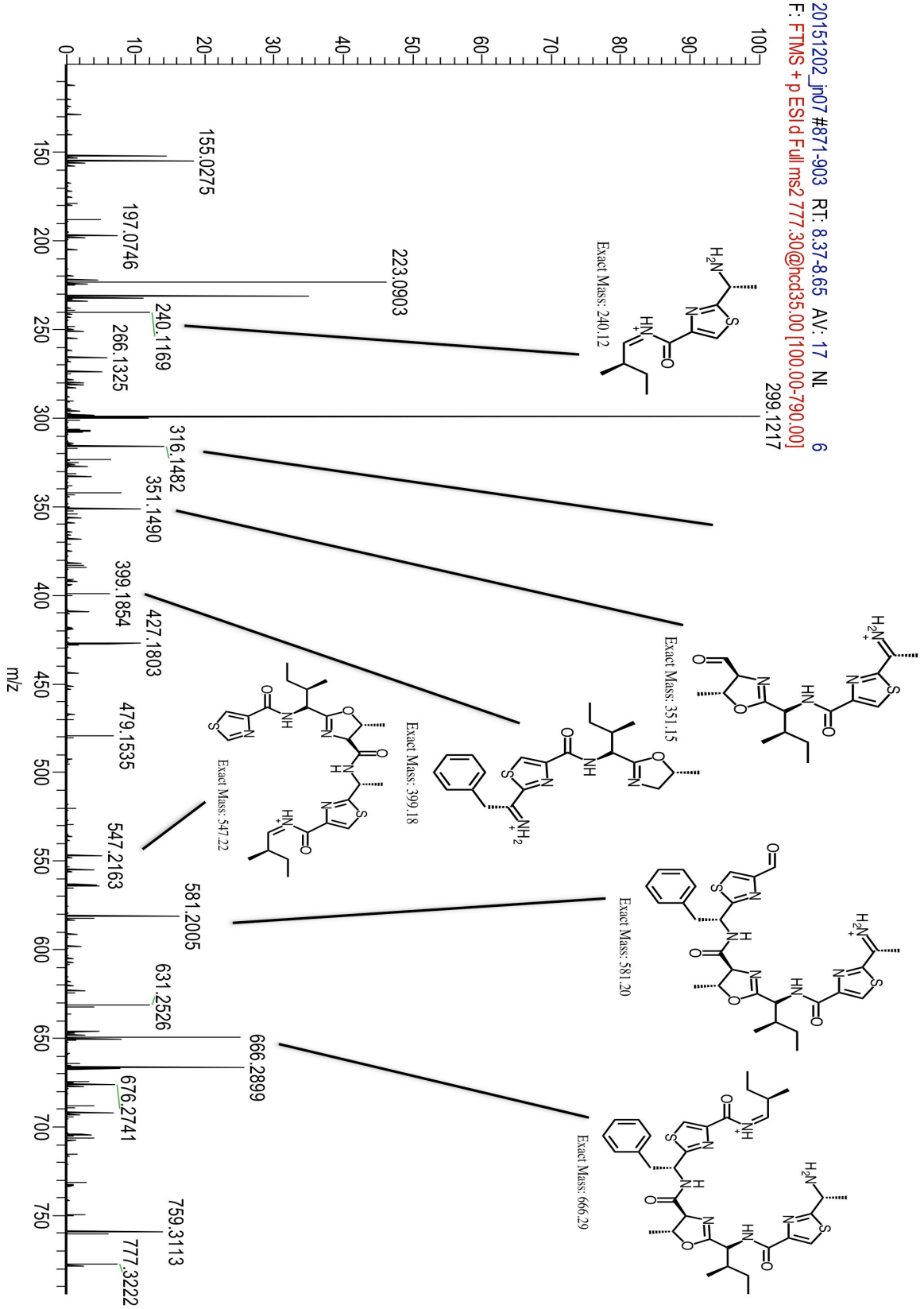



FIGURE 47: **MSMS of cyclo[IT$^{MeOx}$AC$^{Thz}$IT$^{MeOx}$FC$^{Thz}$] (18)**

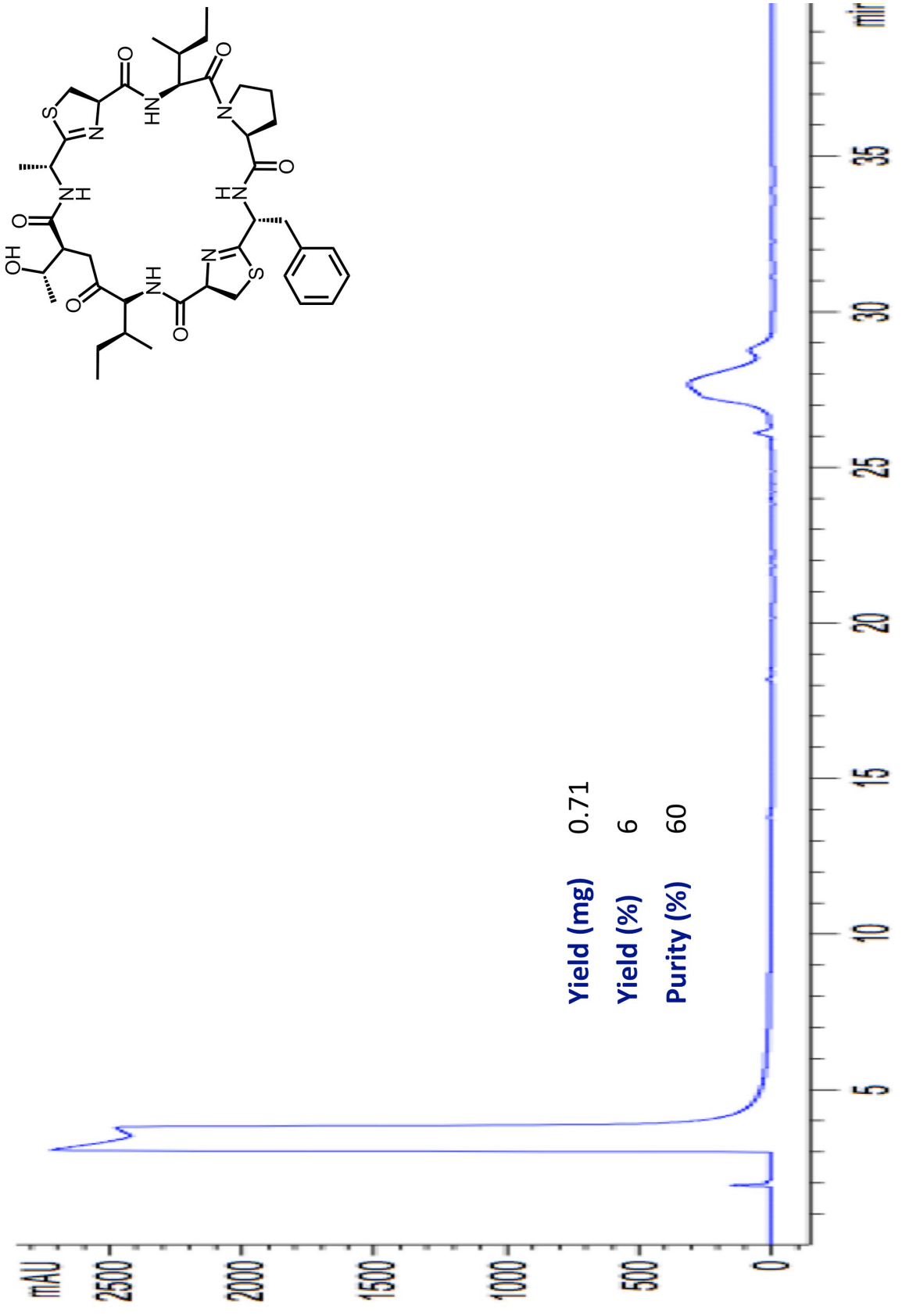



FIGURE 48: **HPLC of cyclo[ITAC$^{ThH}$IPFC$^{ThH}$] (19)**

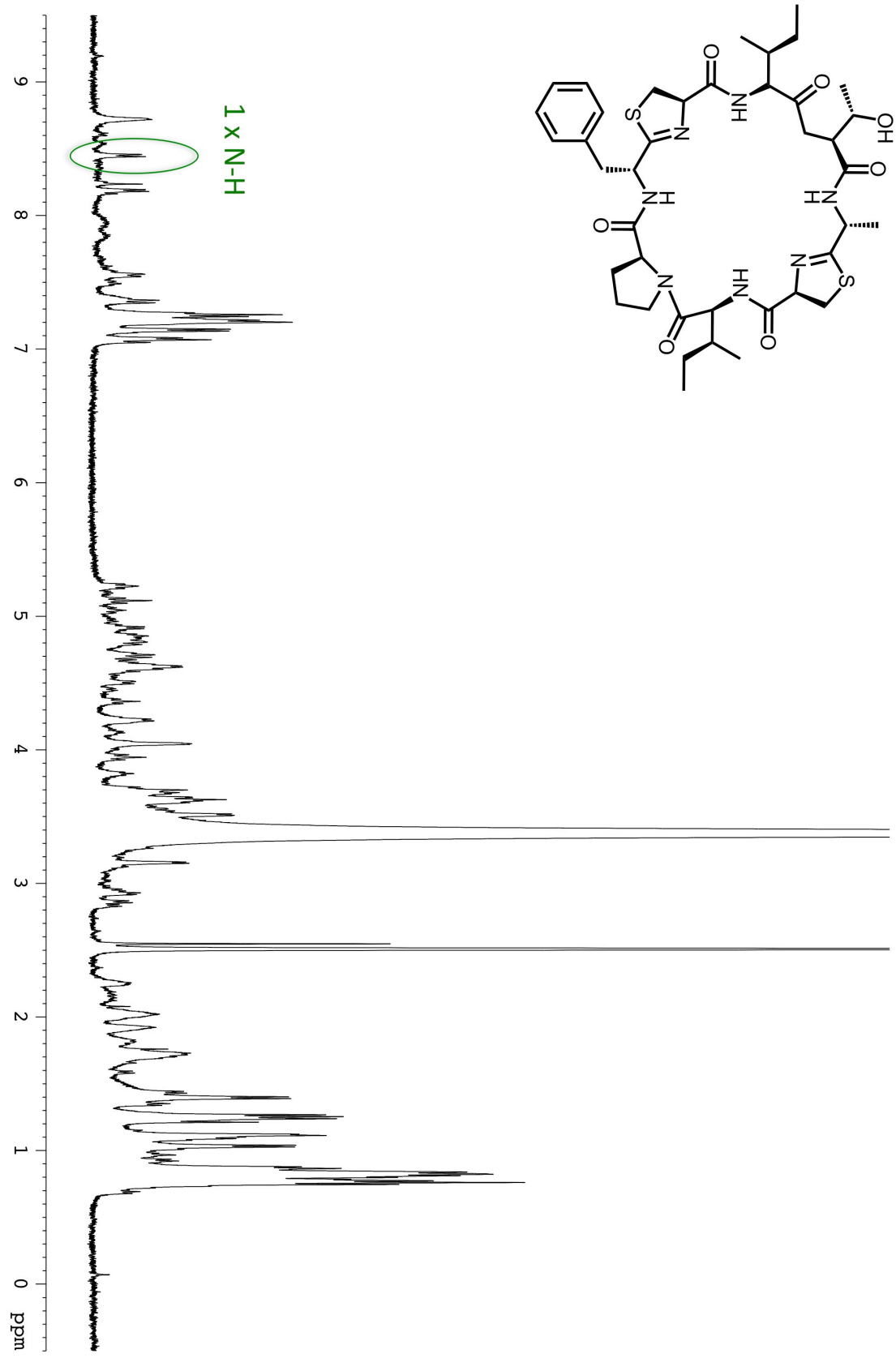



FIGURE 49: **NMR of cyclo[ITAC$^{ThH}$IPFC$^{ThH}$] (19)**

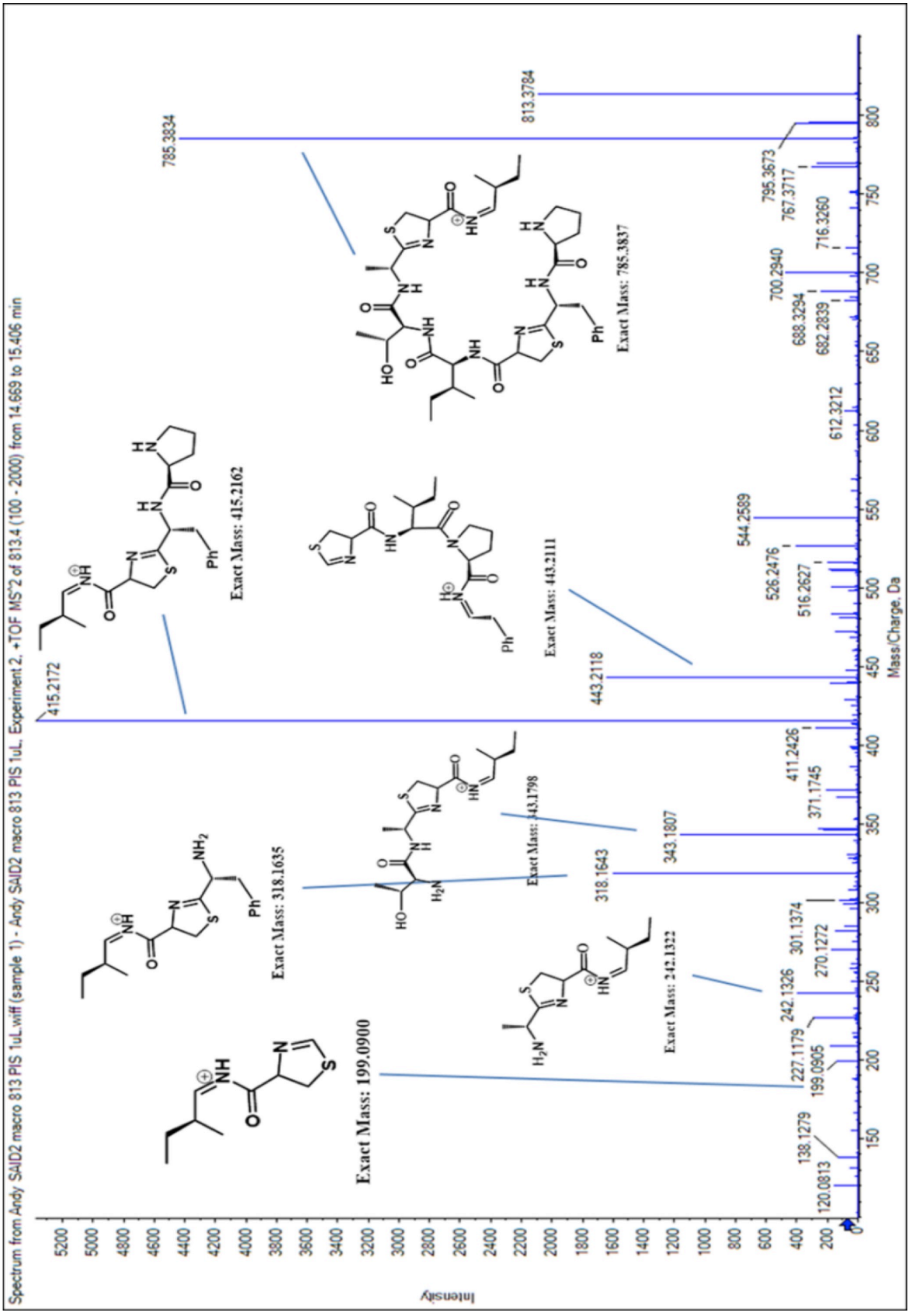



FIGURE 50: **MSMS of cyclo[ITAC$^{\mathrm{ThH}}$IPFC$^{\mathrm{ThH}}$] (19)**

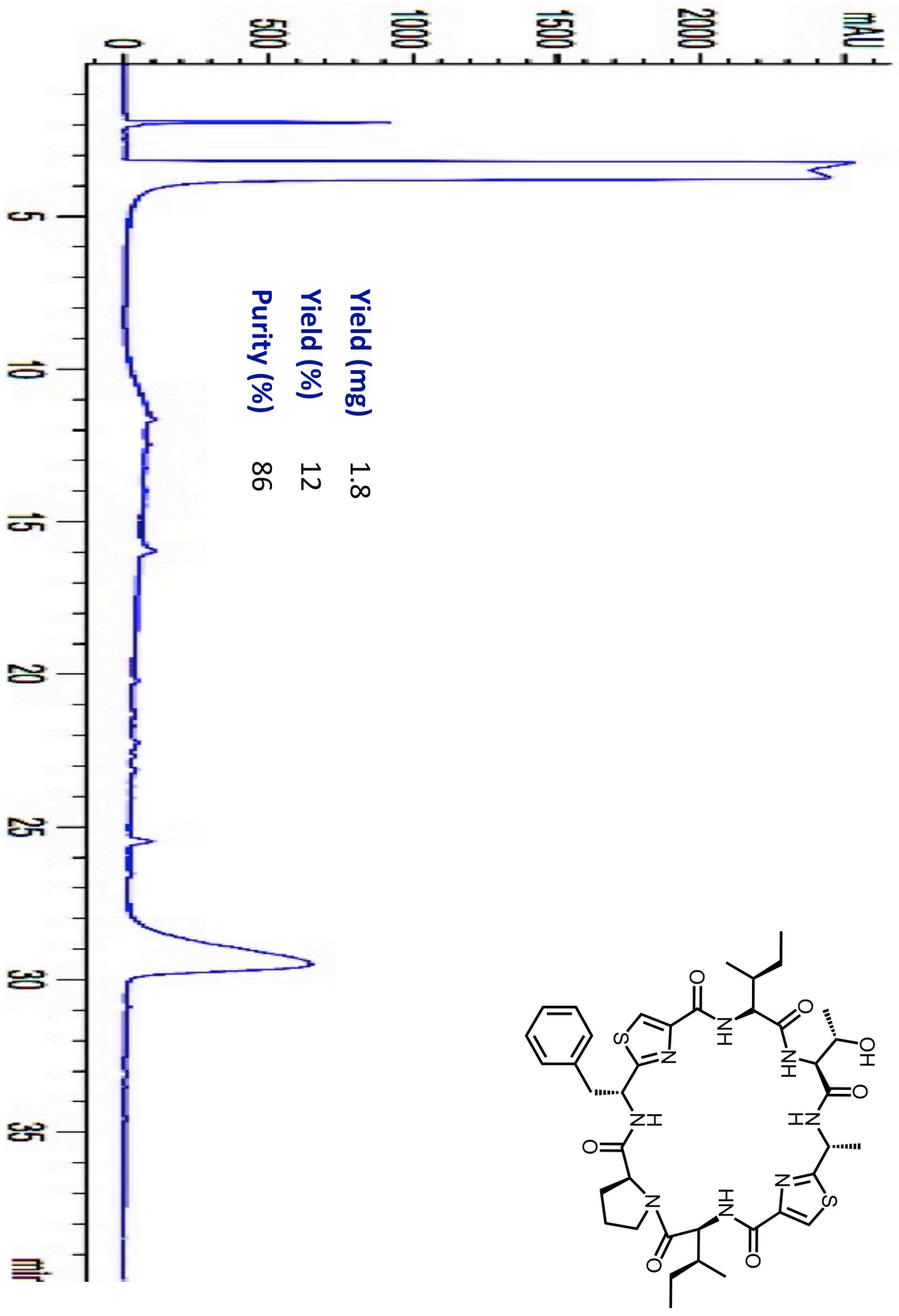



FIGURE 51: **HPLC of cyclo[ITAC$^{Thz}$IPFC$^{Thz}$] (20)**

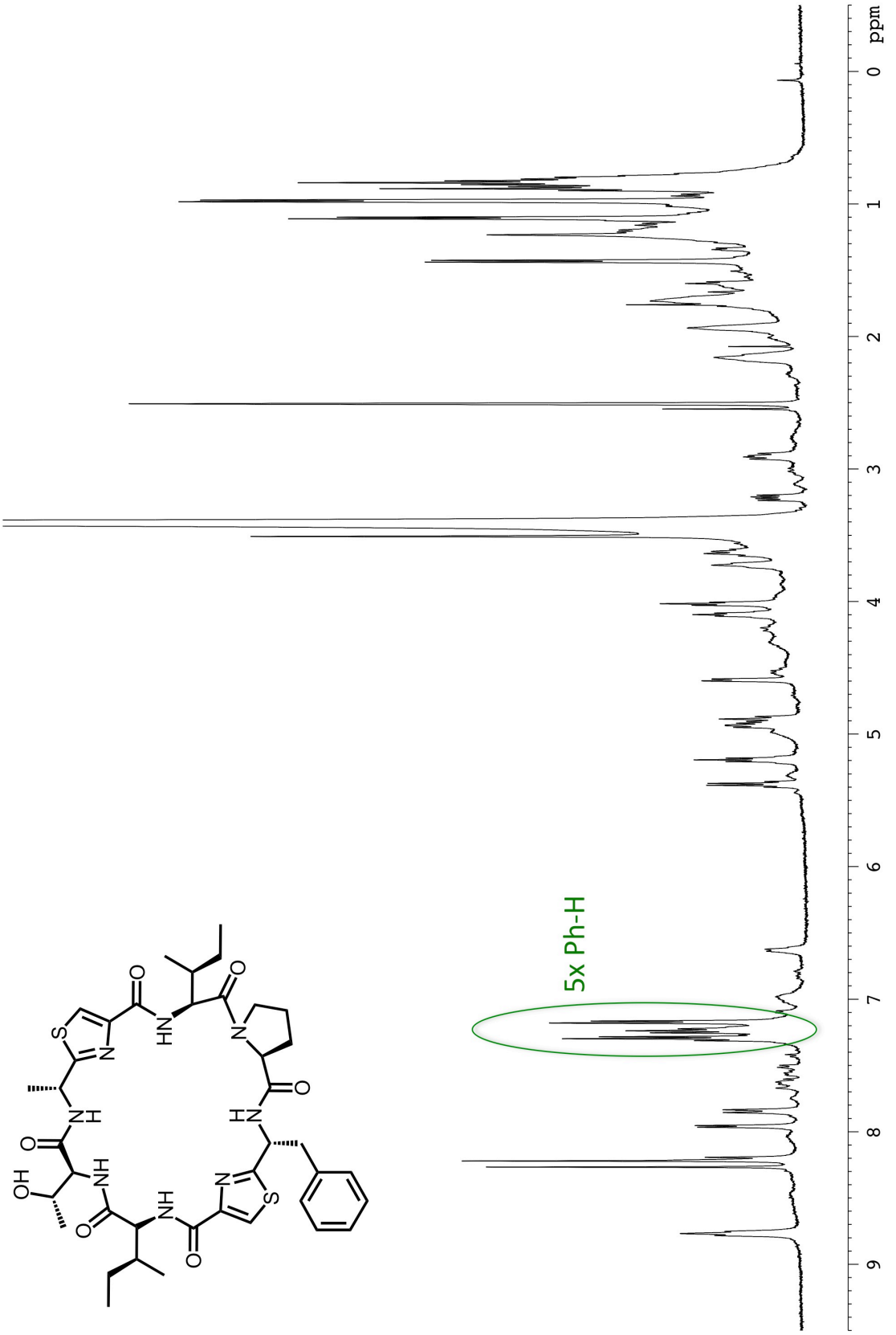



FIGURE 52: **NMR of cyclo[ITAC$^{Thz}$IPFC$^{Thz}$] (20)**

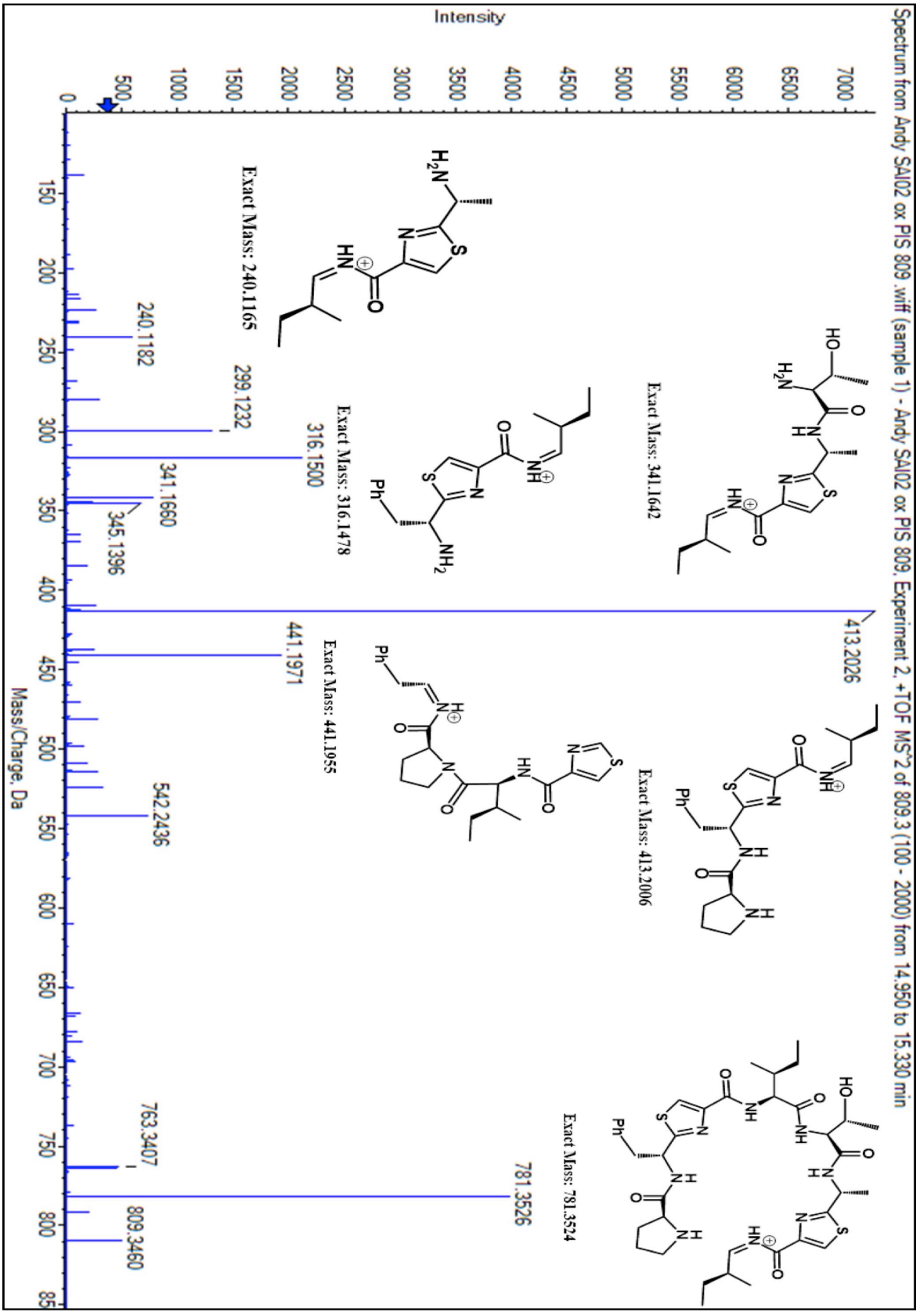



FIGURE 53: **MSMS of cyclo[ITAC$^{Thz}$IPFC$^{Thz}$] (20)**

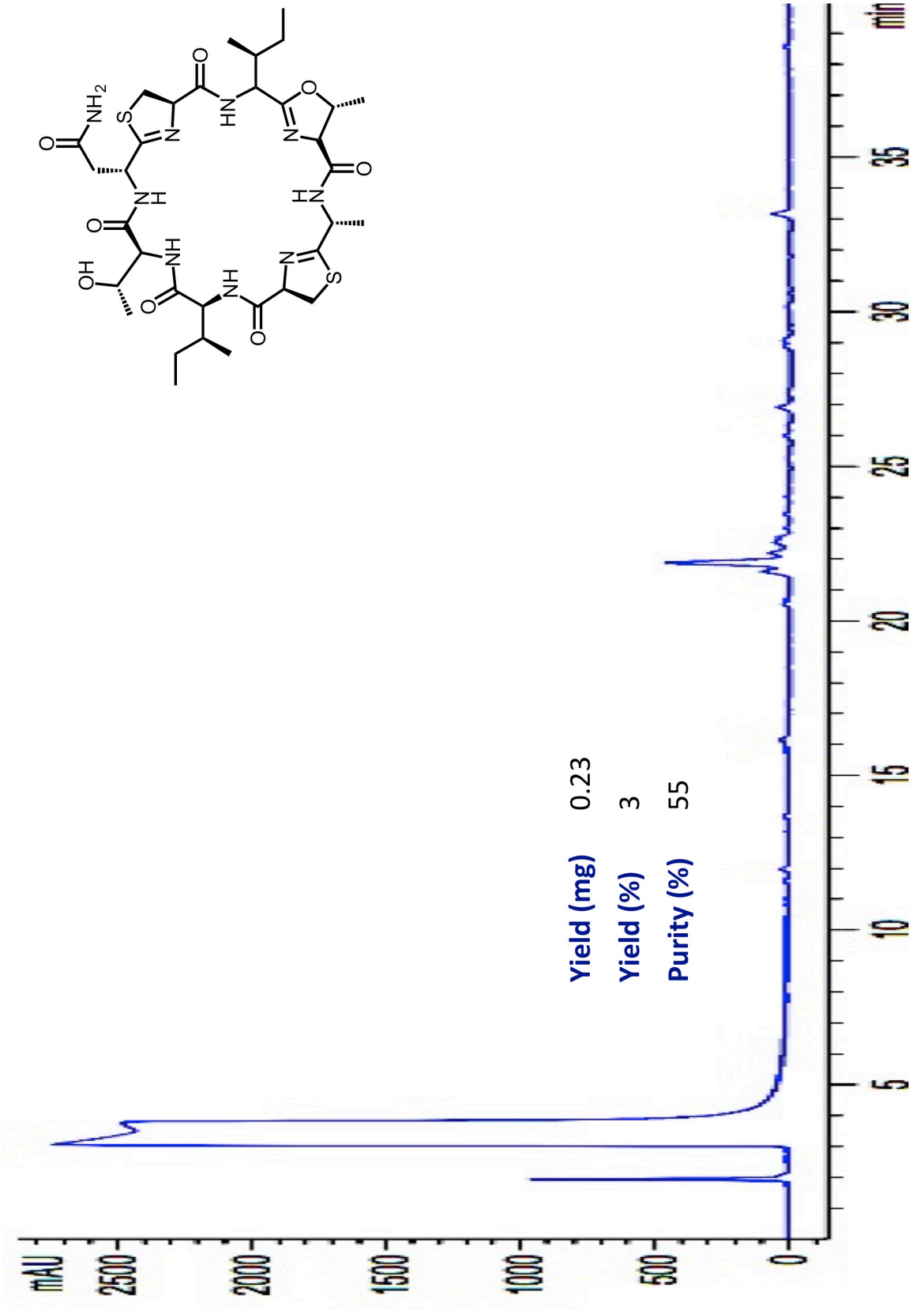



FIGURE 54: **HPLC of cyclo[ITNC$^{ThH}$IT$^{MeOx}$AC$^{ThH}$] (21)**

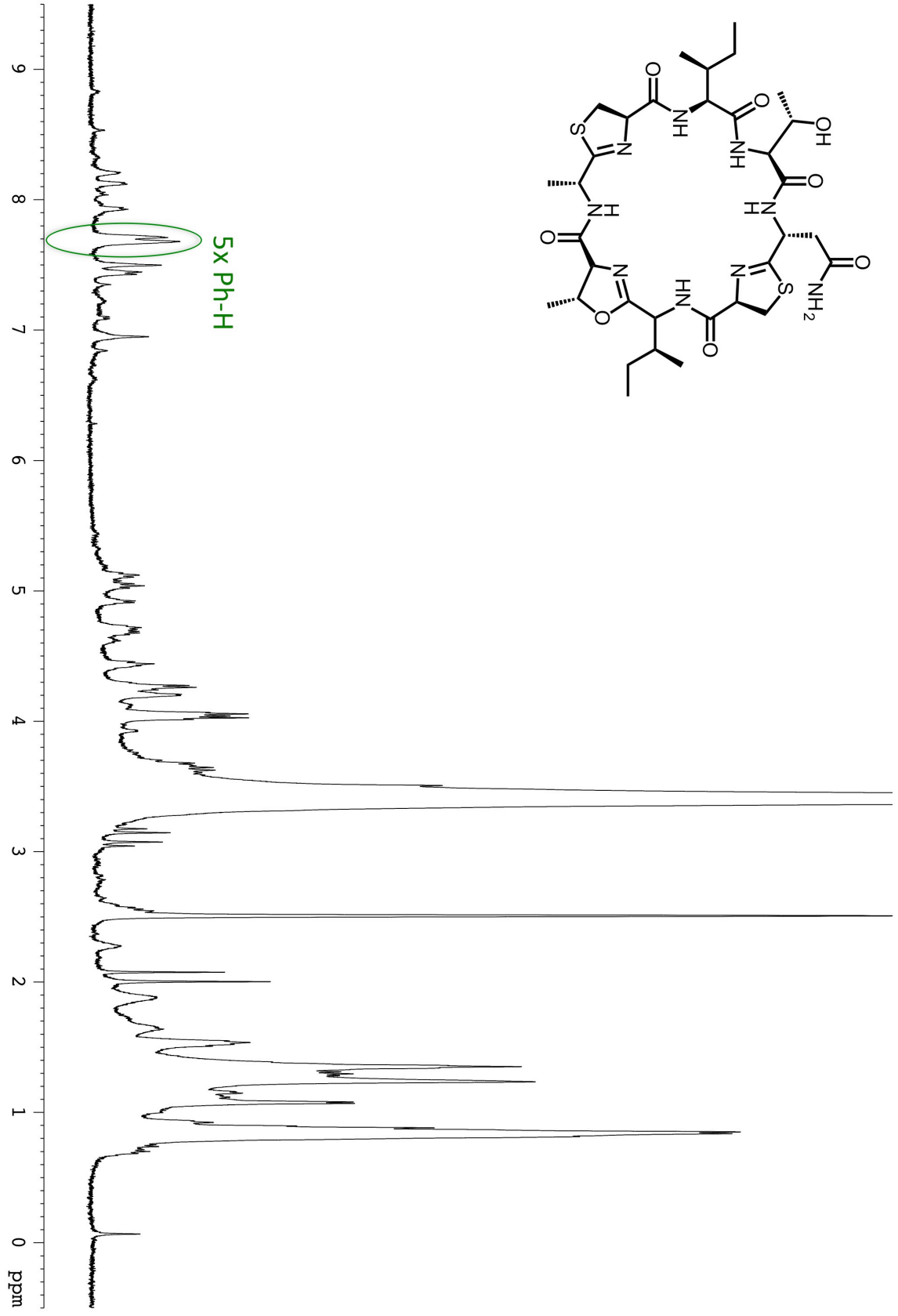



FIGURE 55: **NMR of cyclo[ITNC$^{\text{ThH}}$IT$^{\text{MeOx}}$AC$^{\text{ThH}}$] (21)**

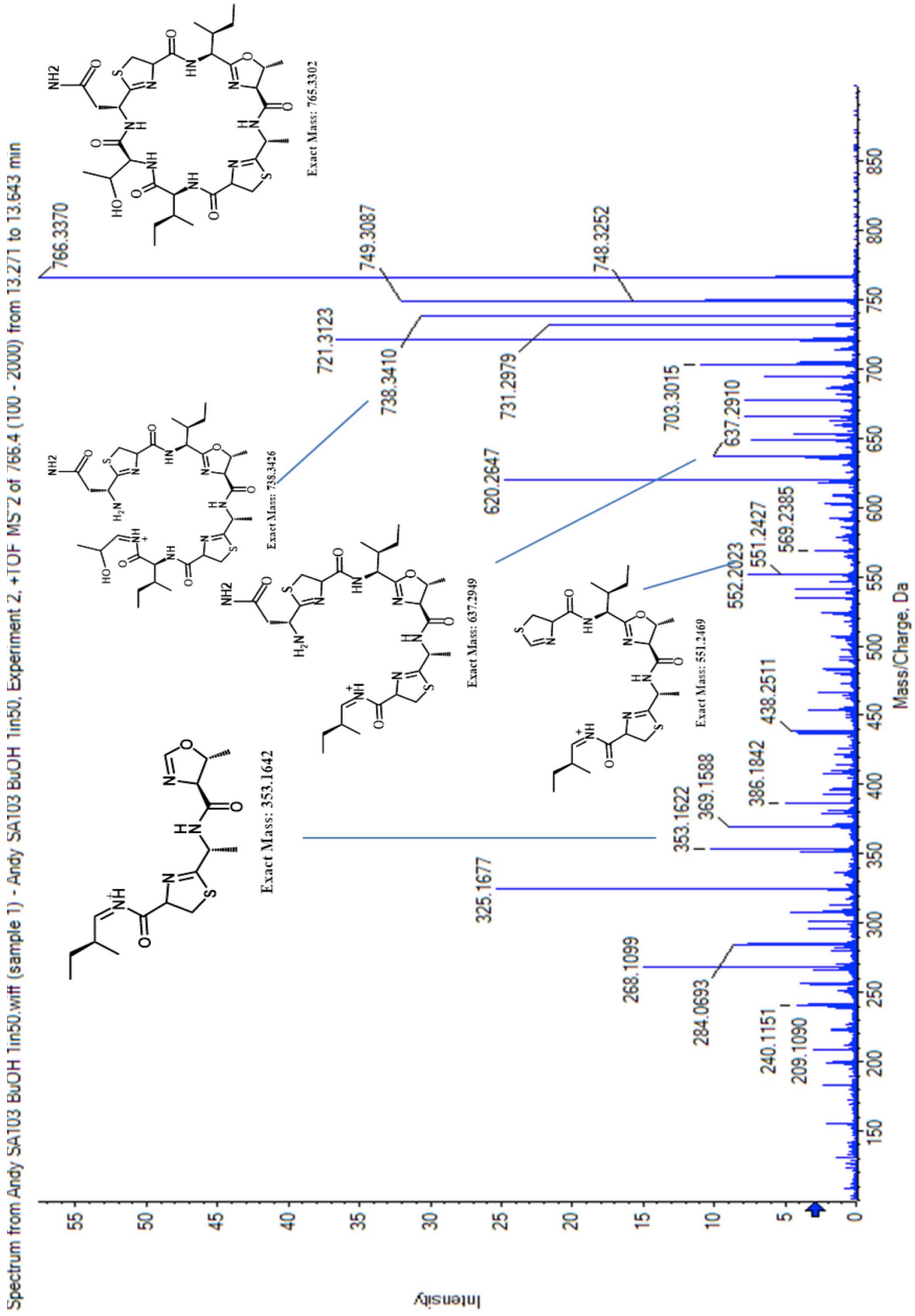



FIGURE 56: **MSMS of cyclo[ITNC$^{ThH}$IT$^{MeOx}$AC$^{ThH}$] (21)**

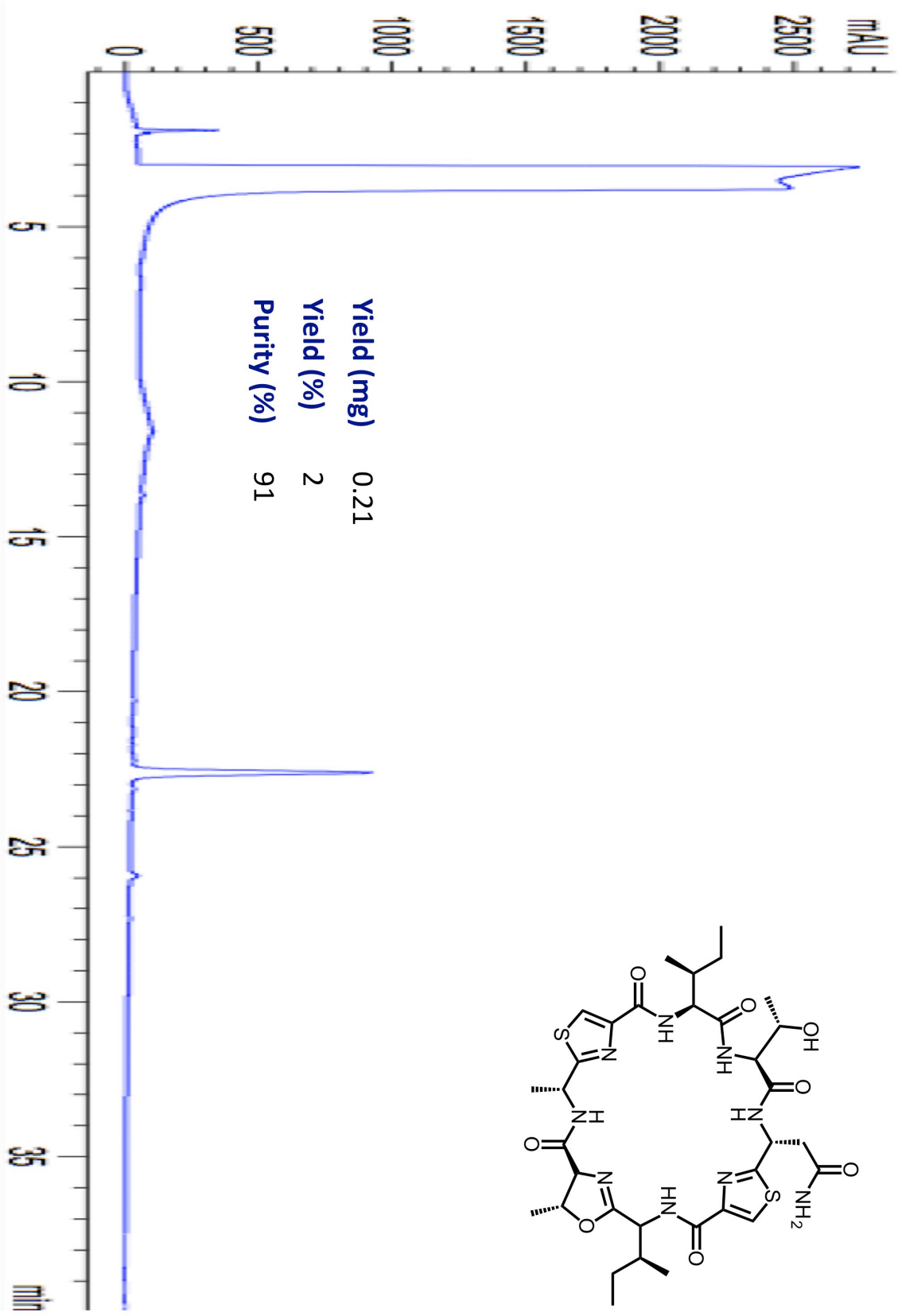



FIGURE 57: **HPLC of cyclo[ITNC$^{\text{ThzI}}$T$^{\text{MeOx}}$AC$^{\text{ThH}}$] (22)**

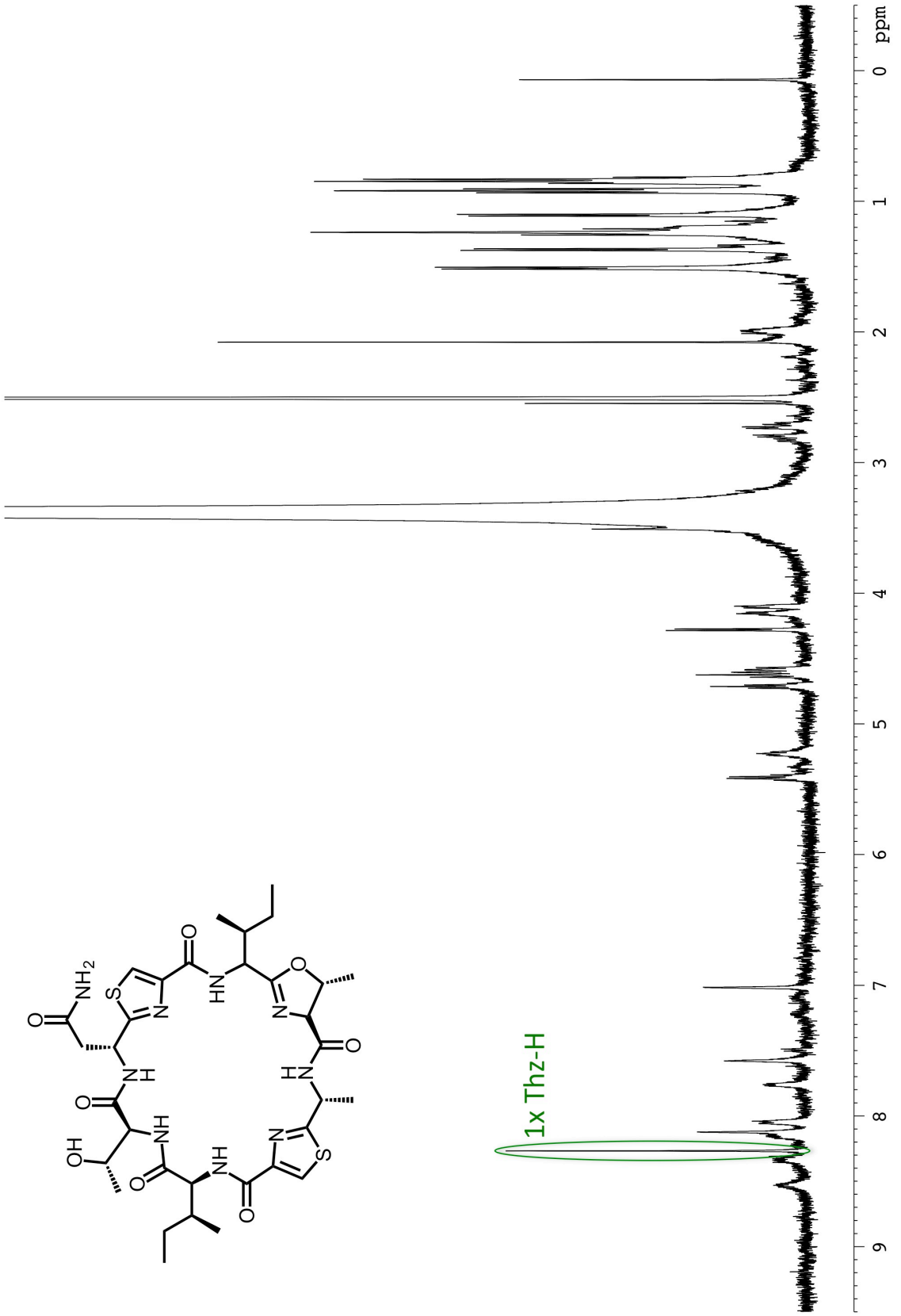



FIGURE 58: **NMR of cyclo[ITNC$^{\mathbf{Thz}}$IT$^{\mathbf{MeOx}}$AC$^{\mathbf{Thz}}$] (22)**

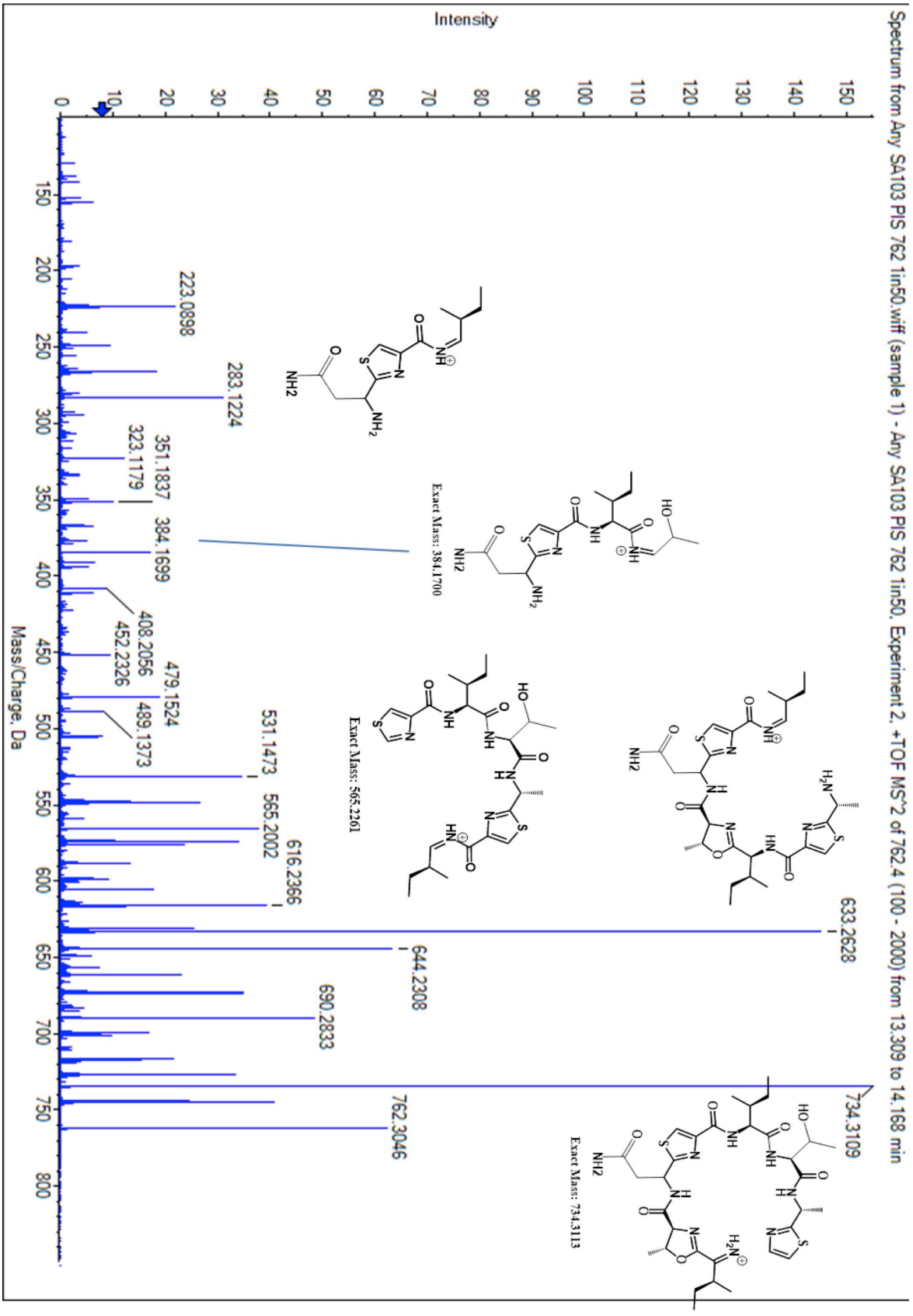



FIGURE 59: **MSMS of cyclo[ITNC$^{Thz}$IT$^{MeOx}$AC$^{Thz}$] (22)**

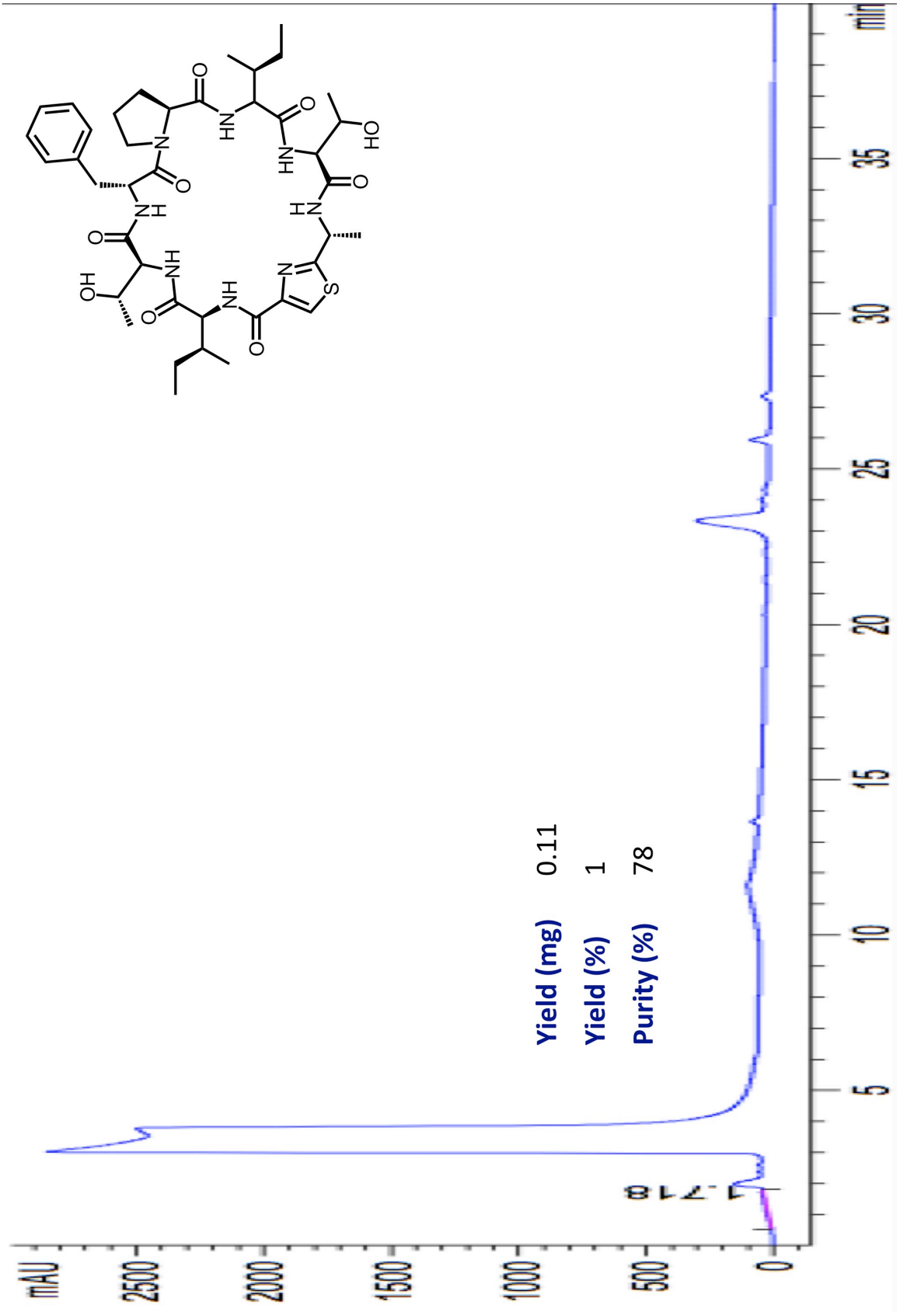



FIGURE 60: **HPLC of cyclo[ITAC$^{Thz}$TFP] (23)**

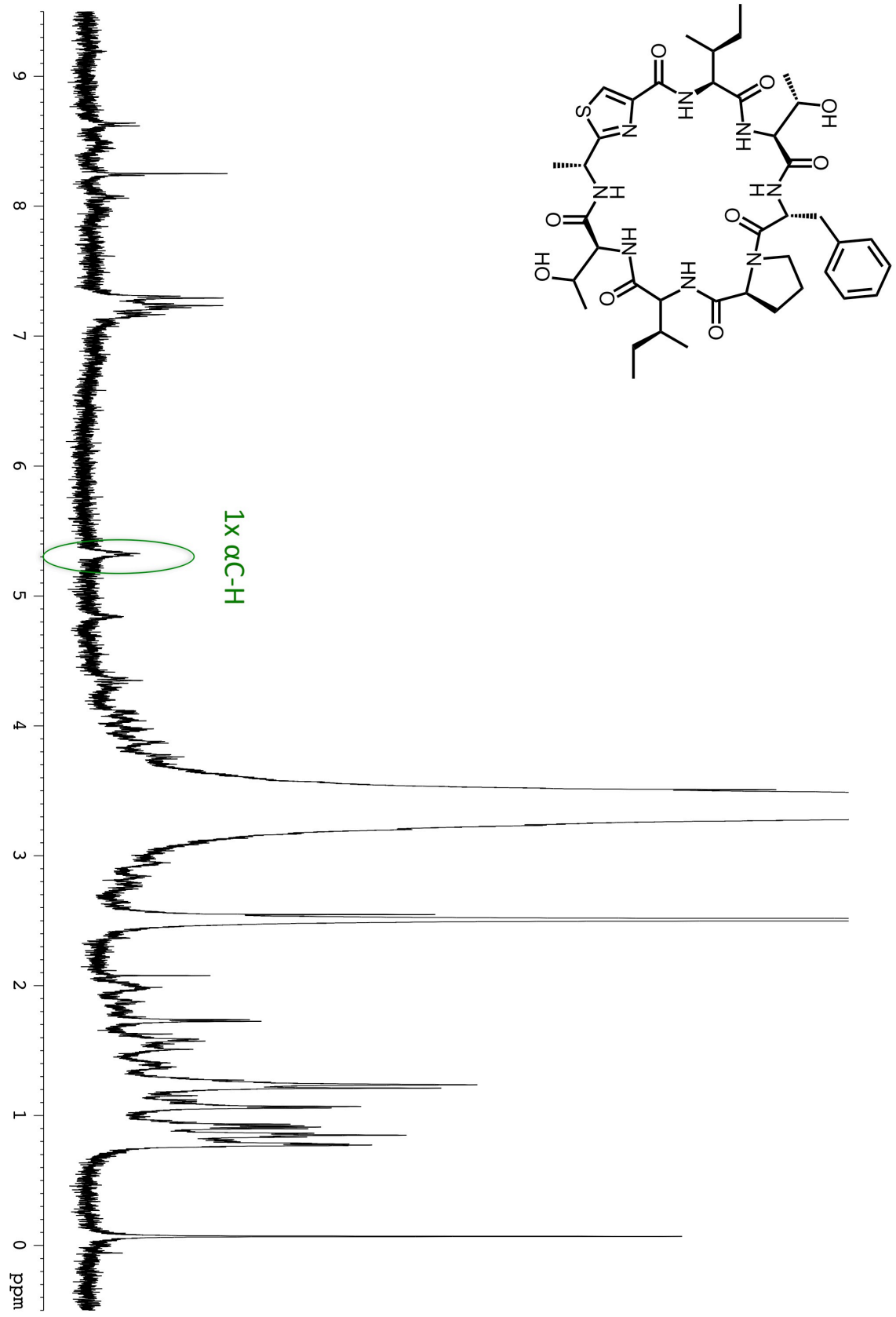



FIGURE 61: **NMR of cyclo[ITAC$^{Thz}$ITFP] (23)**

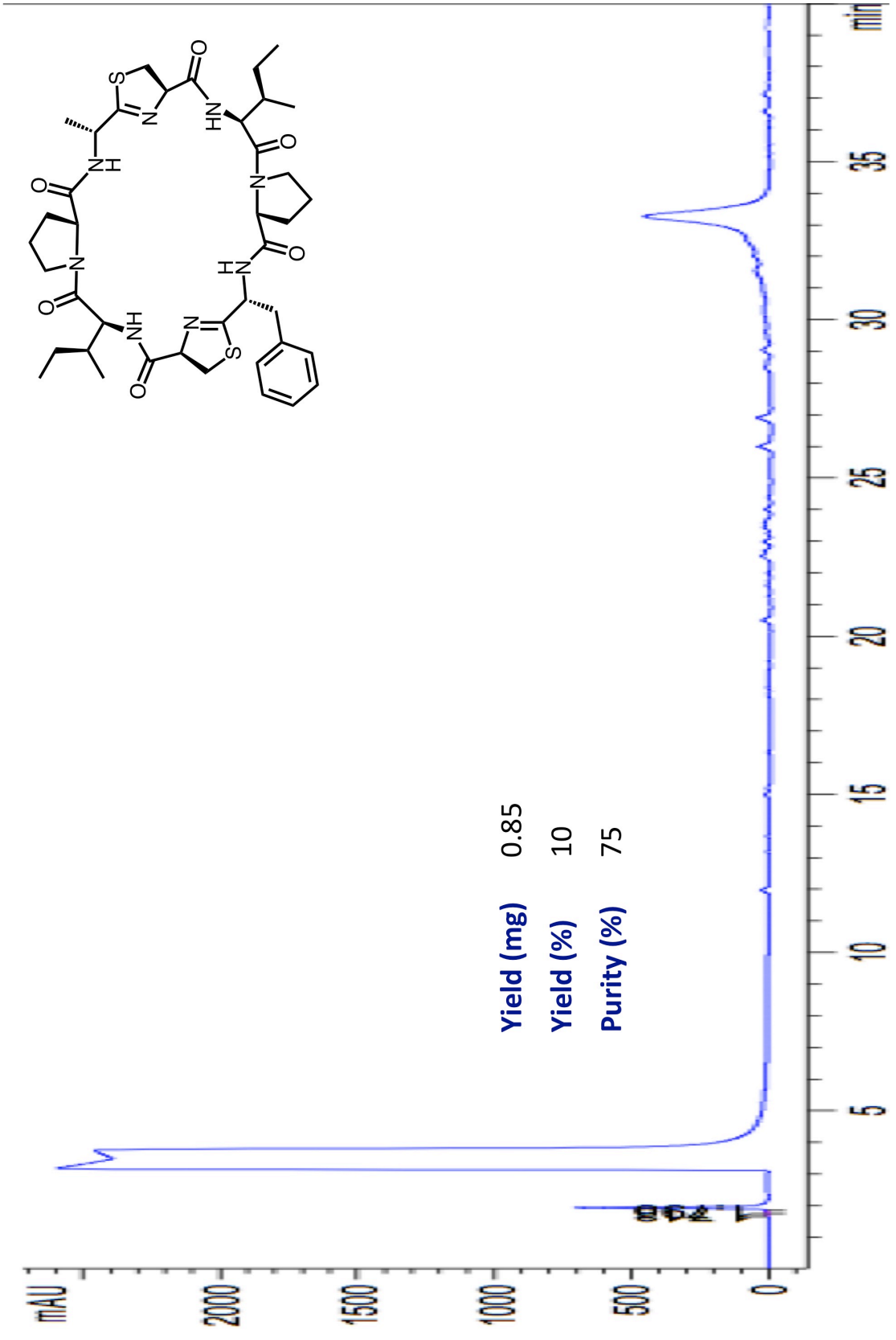



FIGURE 62: **HPLC of cyclo[IPAC$^{ThH}$IPFC$^{ThH}$] (24)**

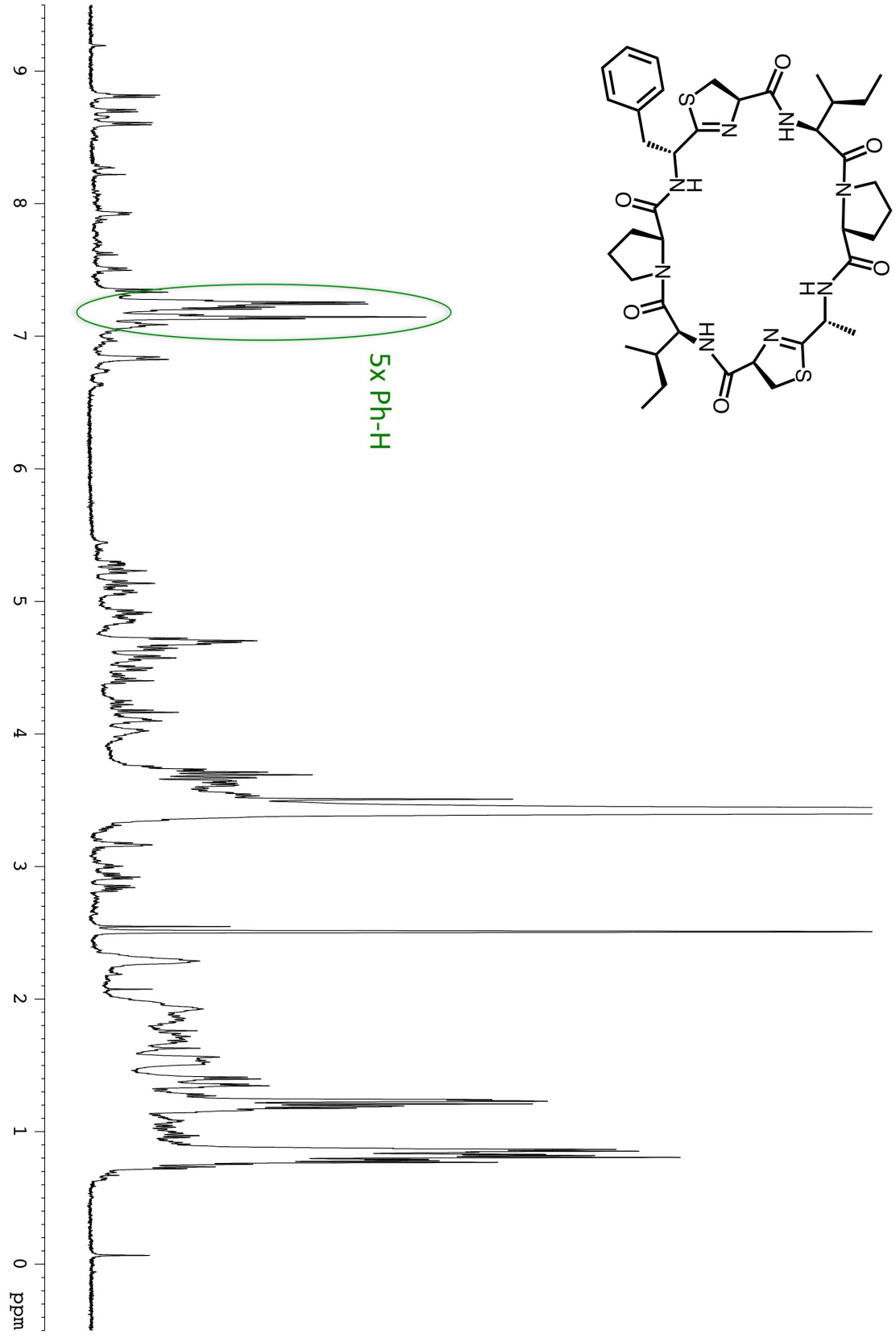



FIGURE 63: **NMR of cyclo[IPAC$^{ThH}$IPFC$^{ThH}$] (24)**

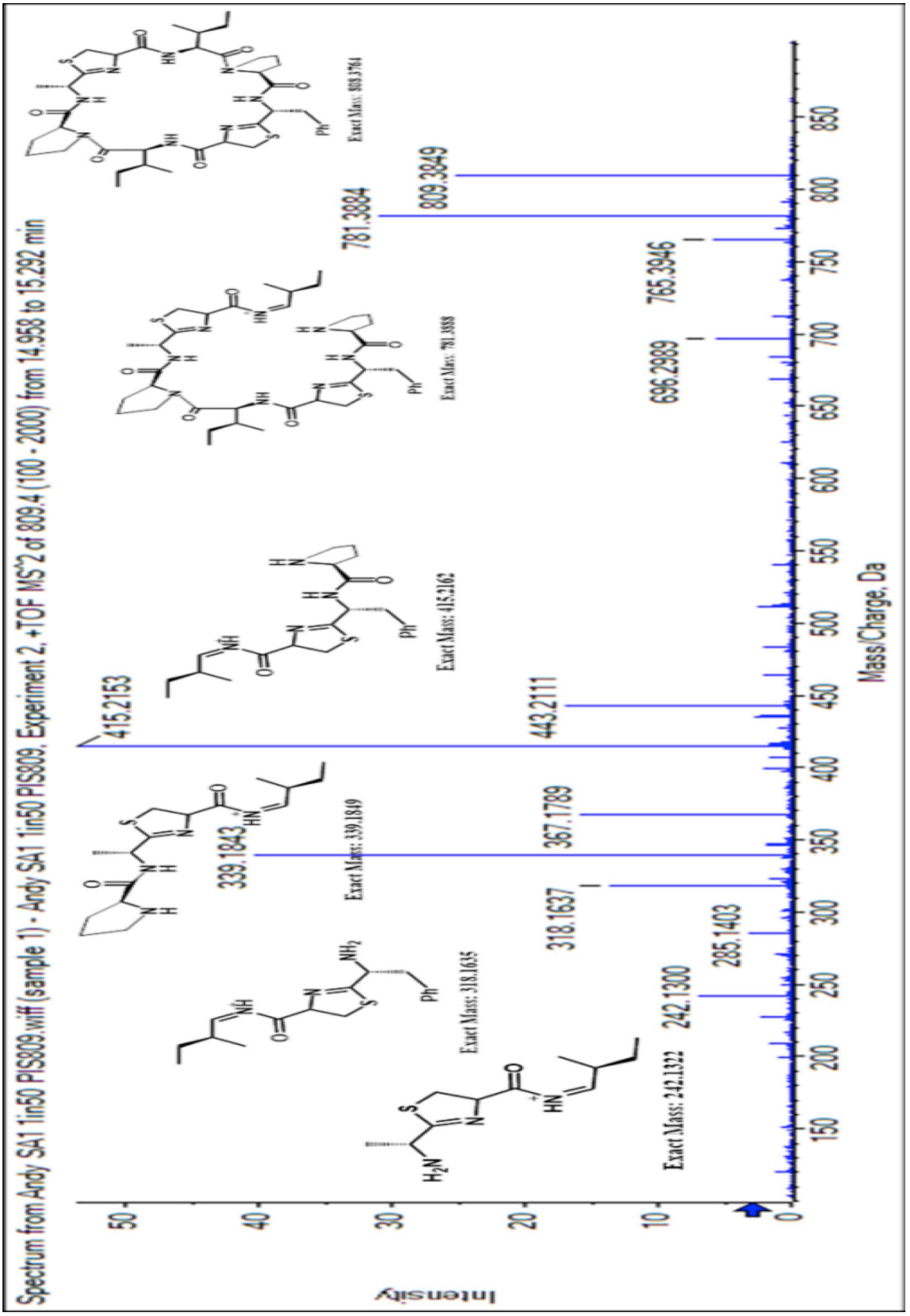



FIGURE 64: **MSMS of cyclo[IPAC$^{ThH}$IPFC$^{ThH}$] (24)**

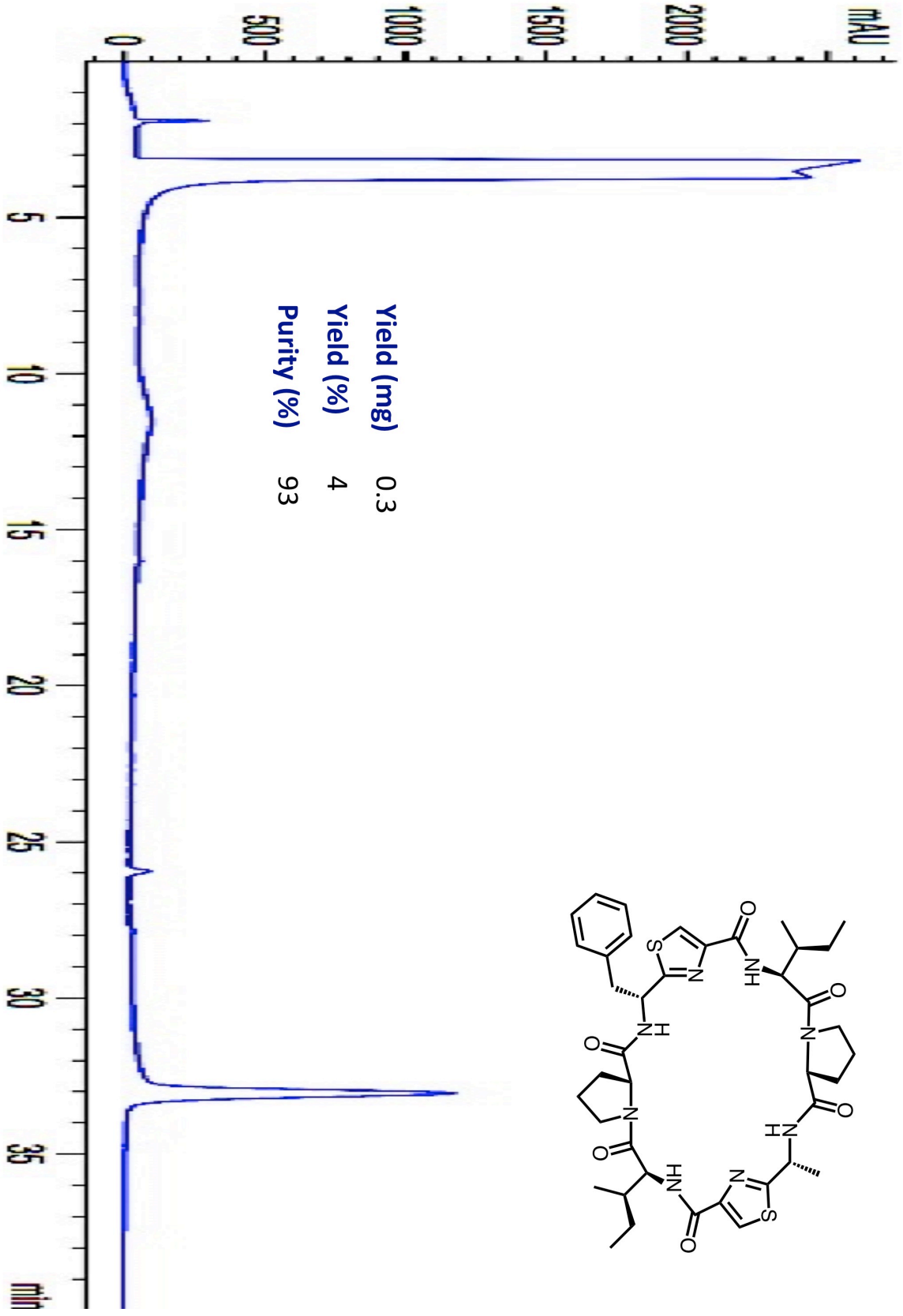



FIGURE 65: **HPLC of cyclo[IPAC$^{Thz}$IPFC$^{Thz}$] (25)**

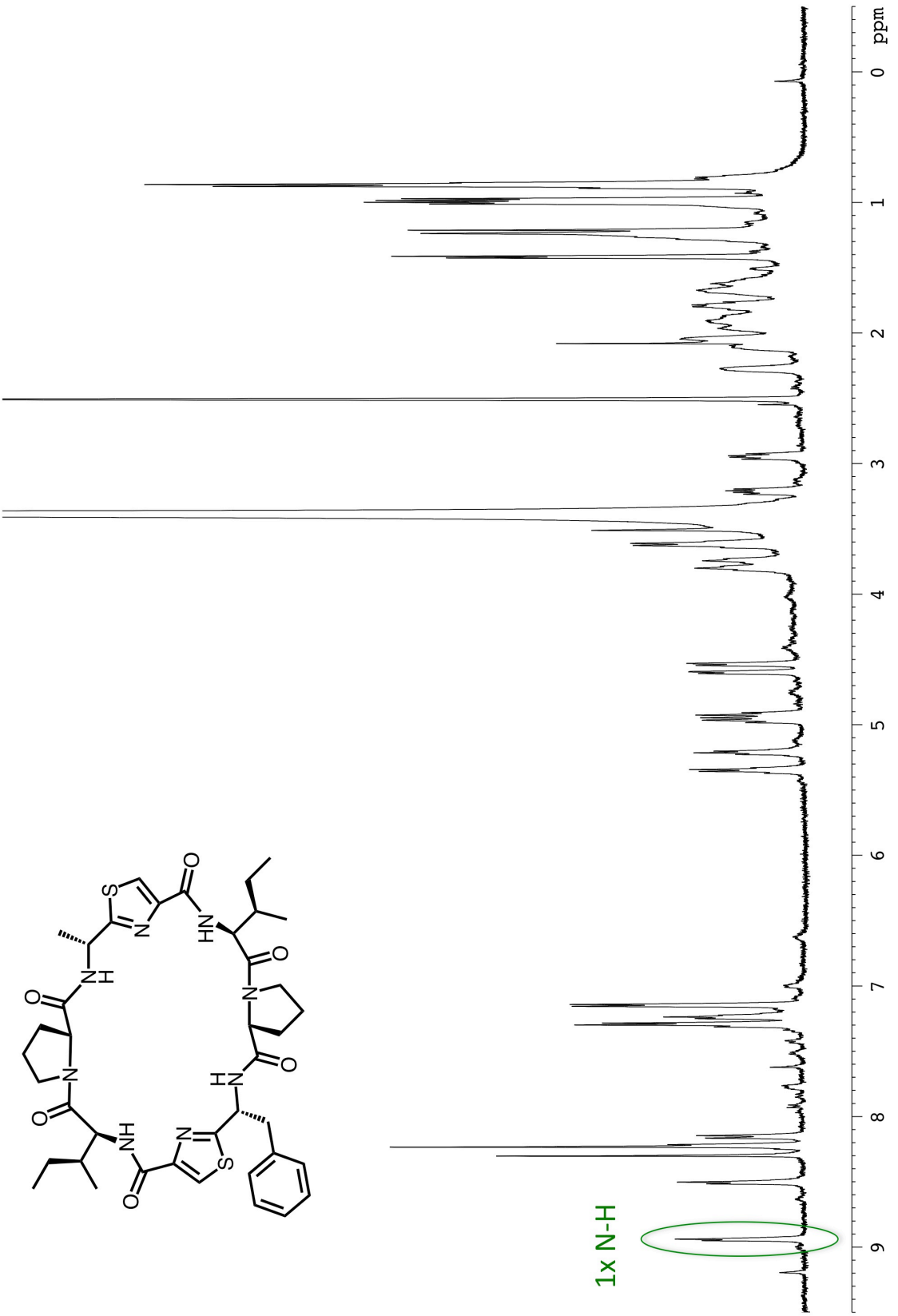



FIGURE 66: **NMR of cyclo[IPAC$^{Thz}$IPFC$^{Thz}$] (25)**

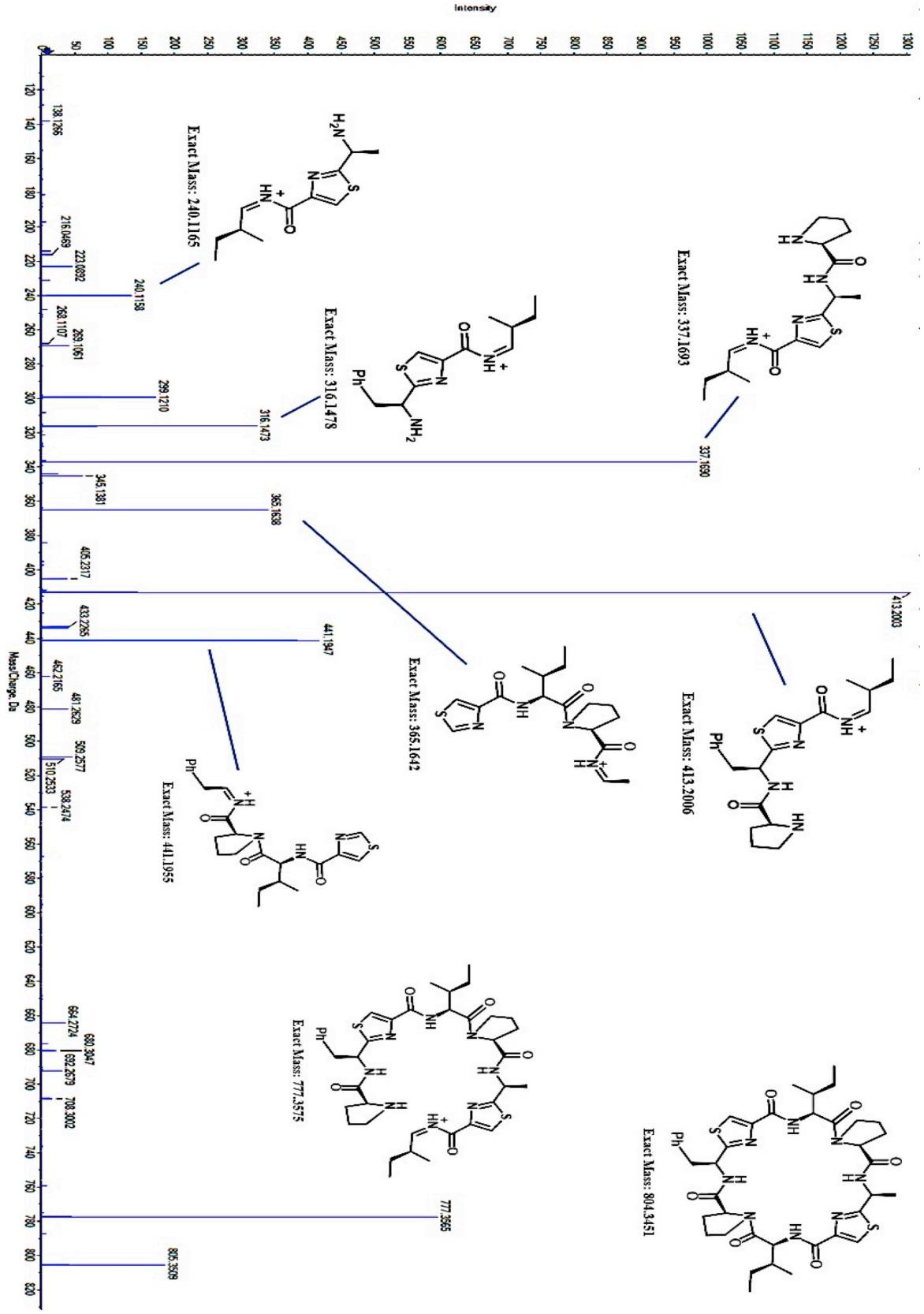



FIGURE 67: **MSMS of cyclo[IPAC$^{Thz}$IPFC$^{Thz}$] (25)**

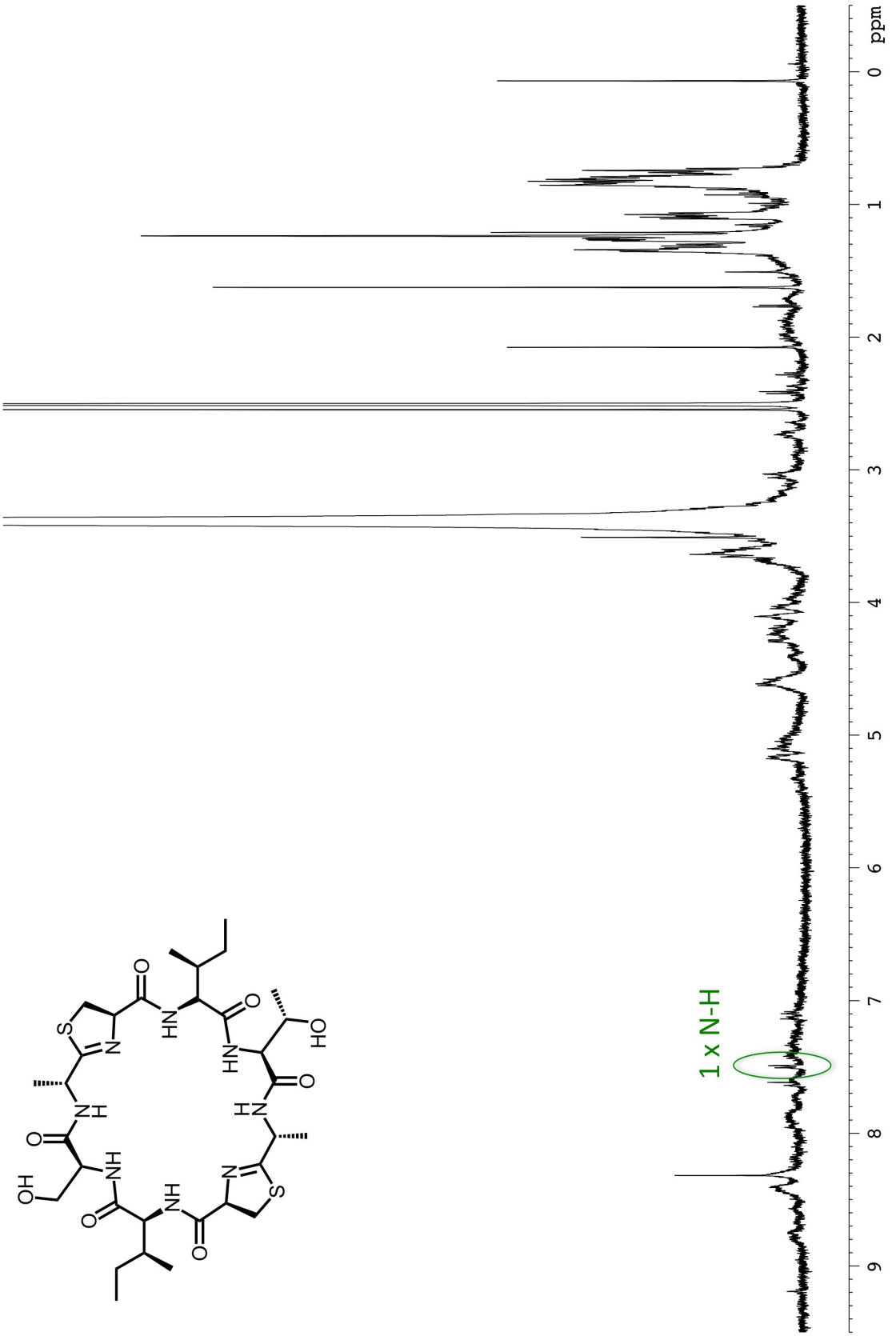



FIGURE 68: **NMR of cyclo[ISAC$^{\mathrm{ThH}}$ITAC$^{\mathrm{ThH}}$] (26)**

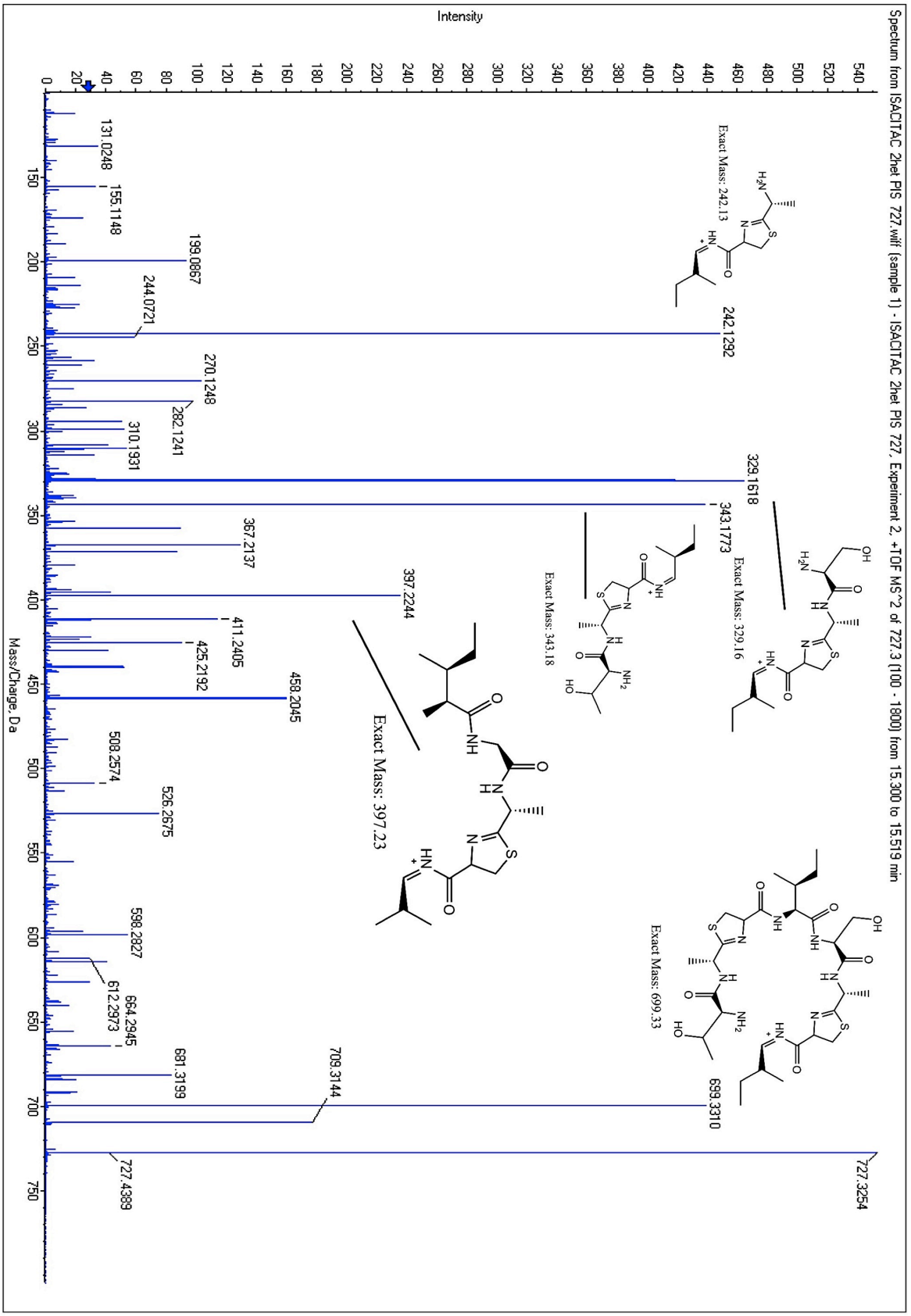



FIGURE 69: **MSMS of cyclo[ISAC$^{\text{ThH}}$ITAC$^{\text{ThH}}$] (26)**

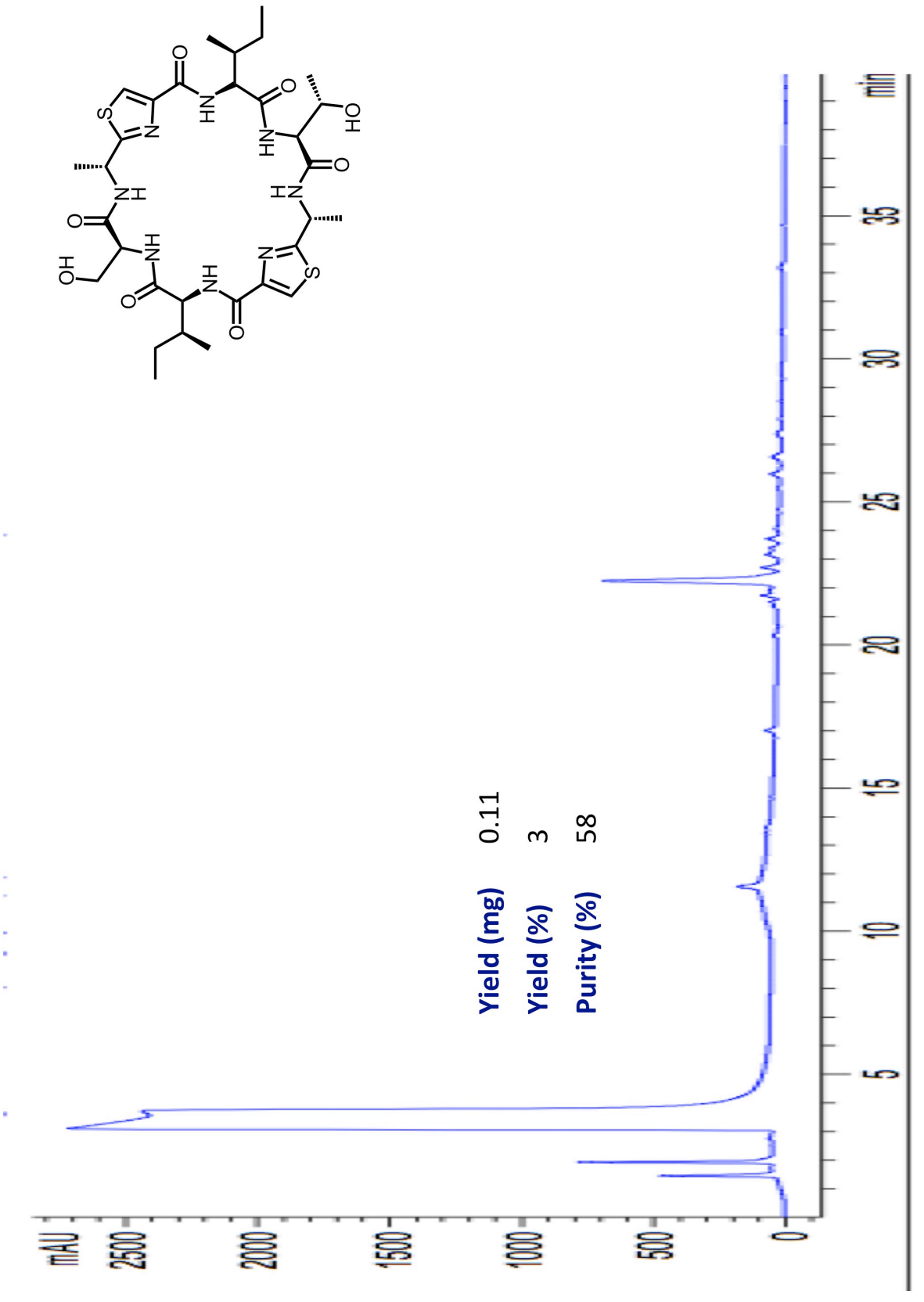



FIGURE 70: **HPLC of cyclo[ISAC$^{Thz}$TAC$^{Thz}$] (27)**

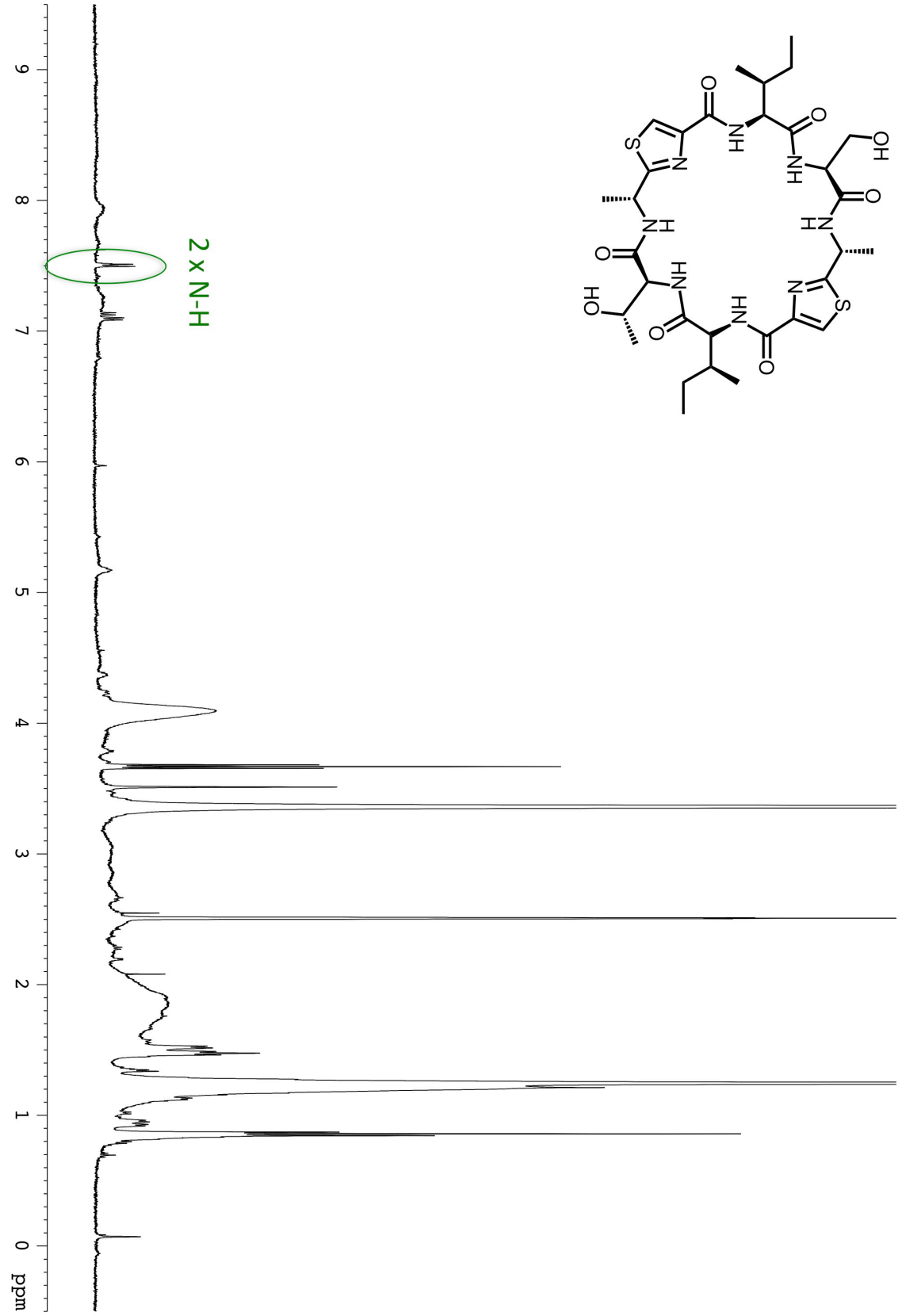



FIGURE 71: **NMR of cyclo[ISAC$^{Thz}$ITAC$^{Thz}$] (27)**

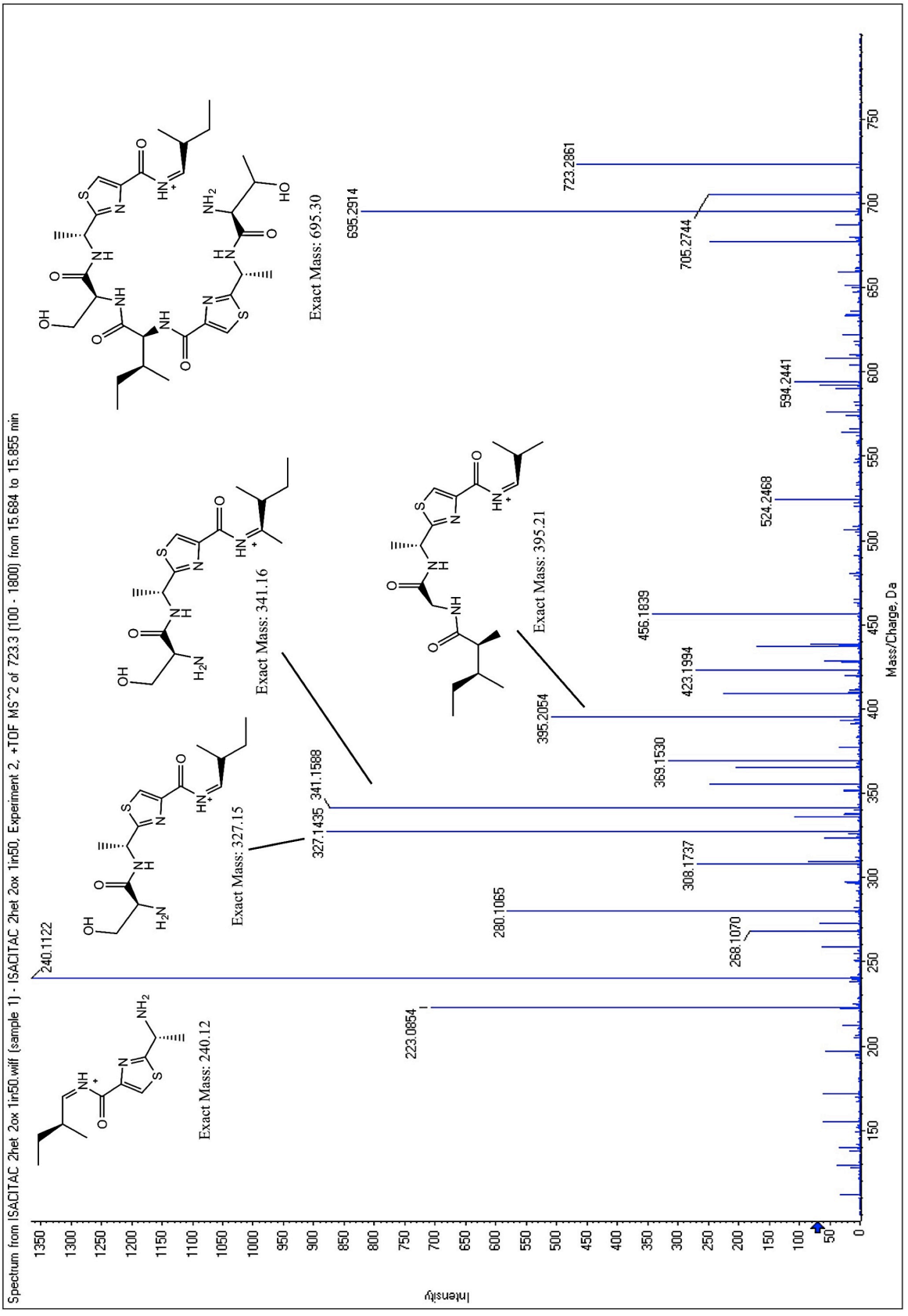



FIGURE 72: **MSMS of cyclo[ISAC$^{\mathbf{Thz}}$ITAC$^{\mathbf{Thz}}$] (27)**

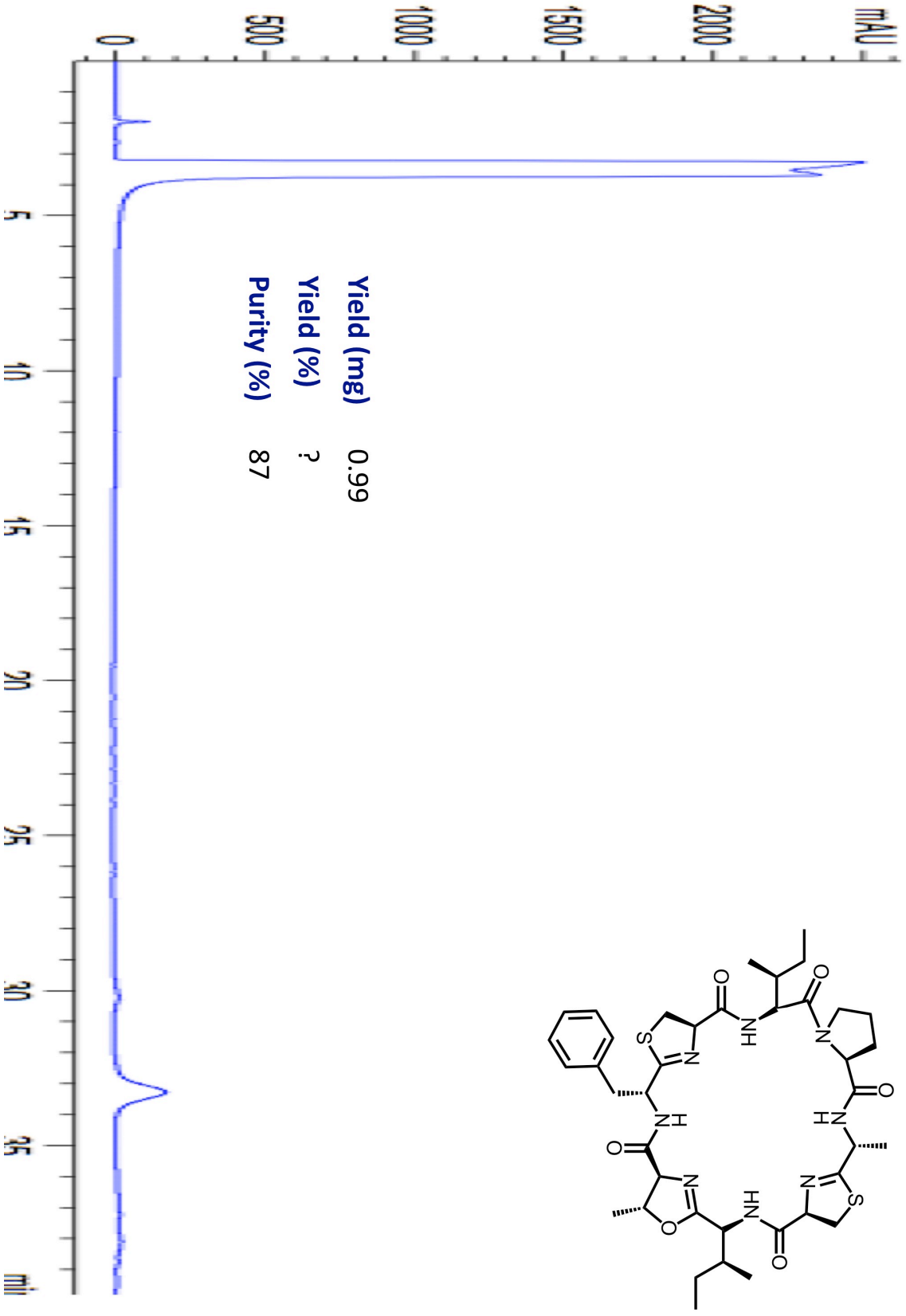



FIGURE 73: **HPLC of cyclo[IPAC$^{\text{ThH}}$IT$^{\text{MeO×F}}$C$^{\text{ThH}}$] (28)**

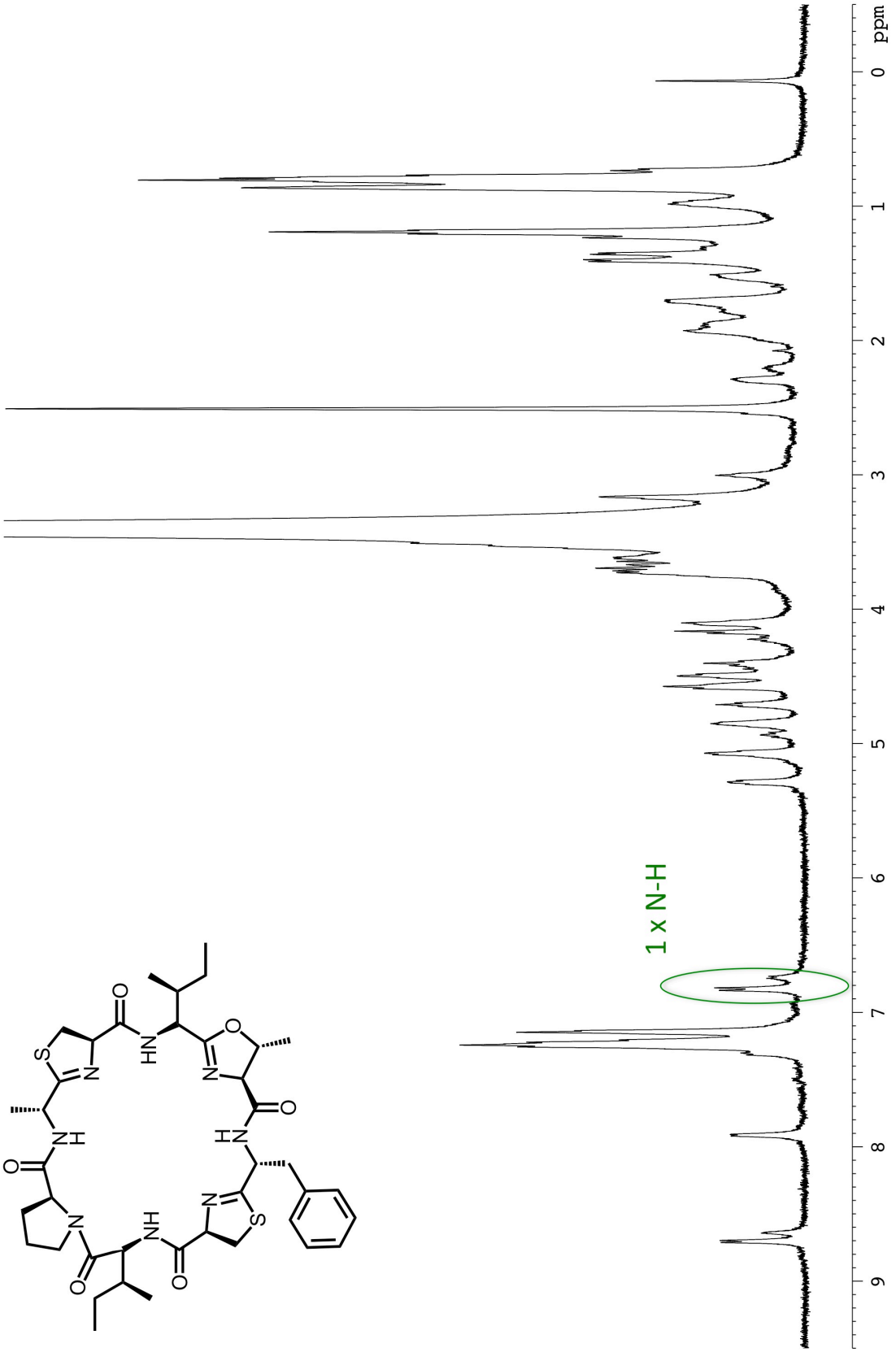



FIGURE 74: **NMR of cyclo[IPAC$^{\mathbf{ThH}}$IT$^{\mathbf{MeOx}}$FC$^{\mathbf{ThH}}$] (28)**

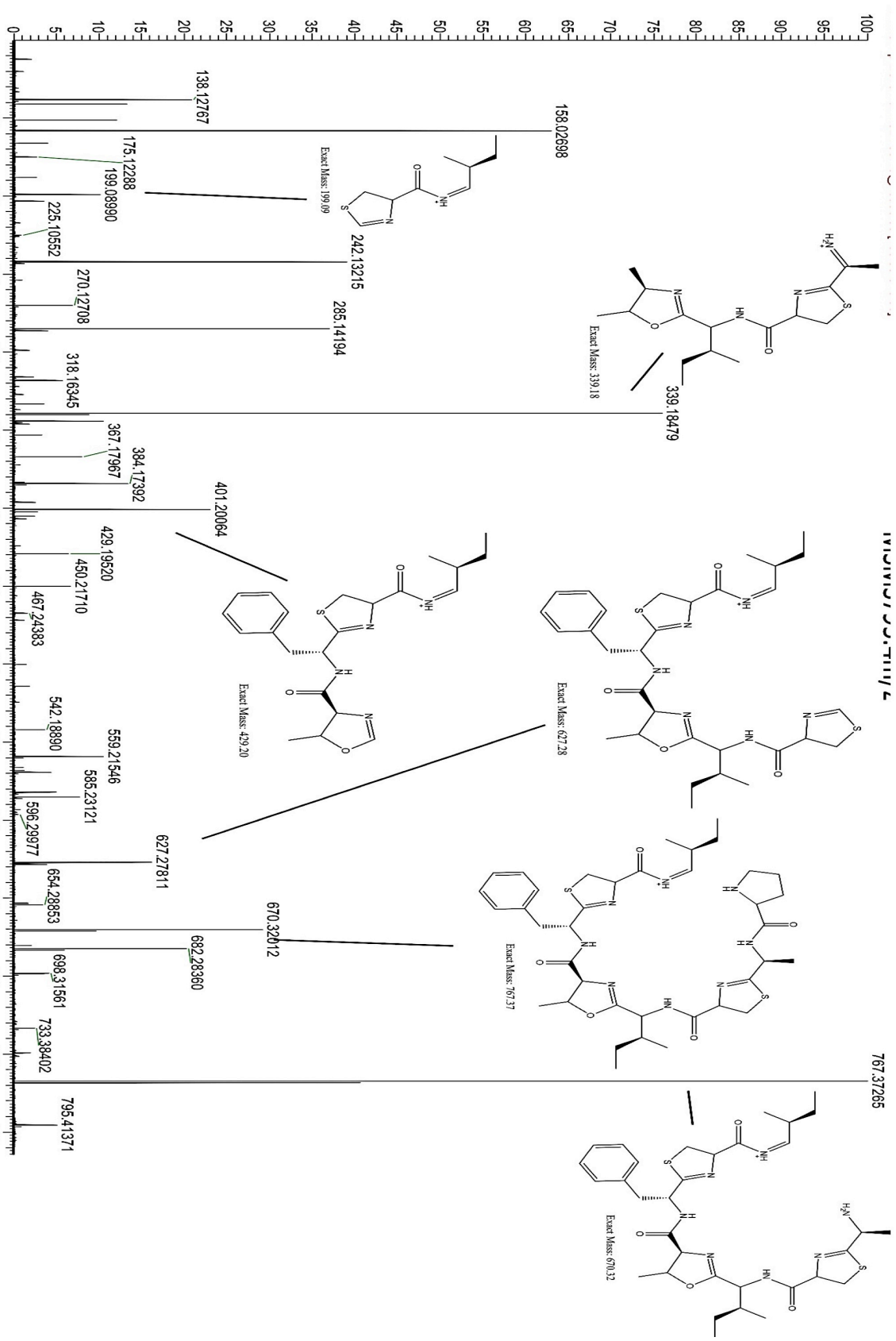



FIGURE 75: **MSMS of cyclo[IPAC$^{\mathbf{ThH}}$IT$^{\mathbf{MeO\times}}$FC$^{\mathbf{ThH}}$] (28)**

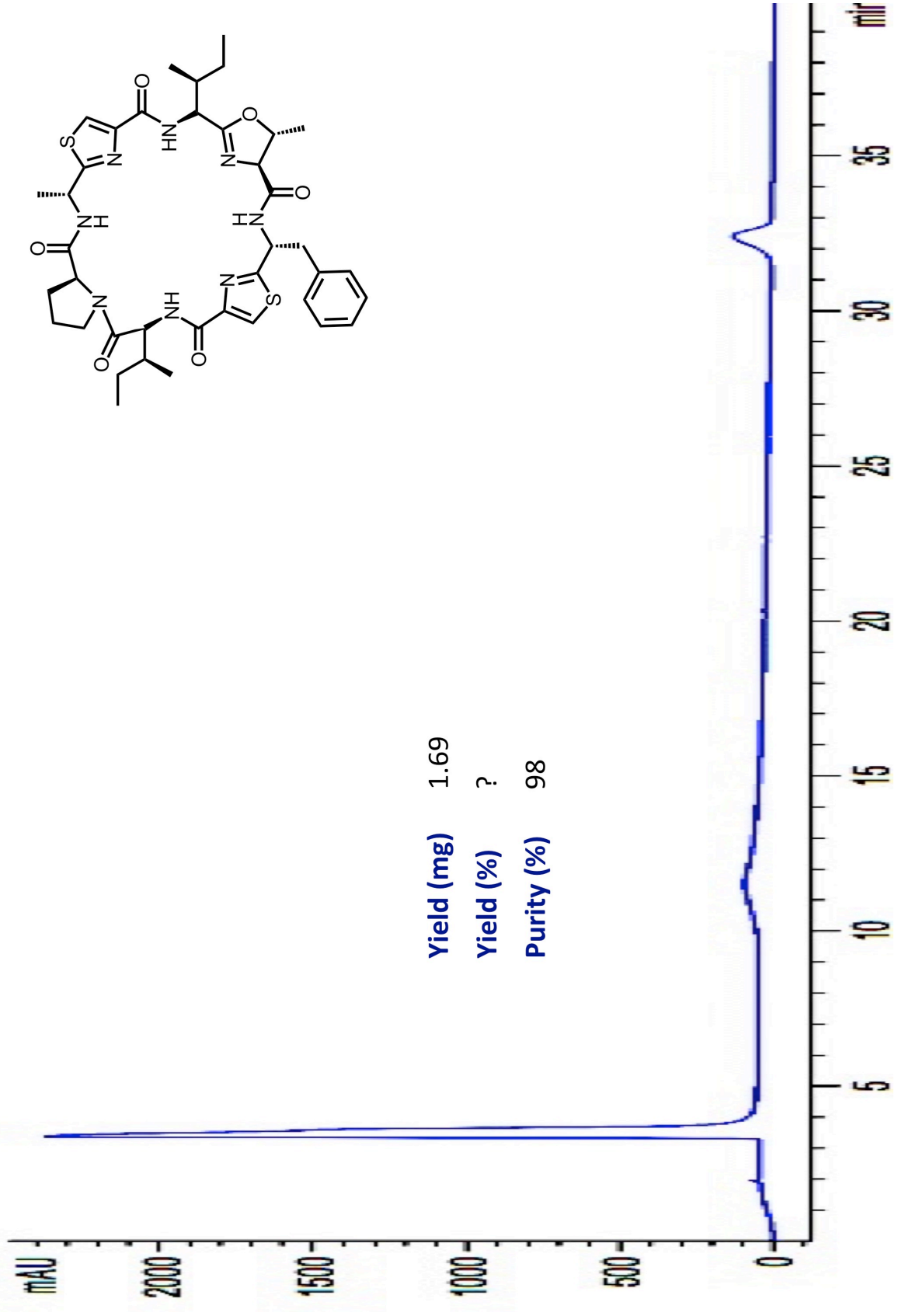



FIGURE 76: **HPLC of cyclo[IPAC$^{\mathbf{Thz}}$IT$^{\mathbf{MeO\times}}$FC$^{\mathbf{Thz}}$] (29)**

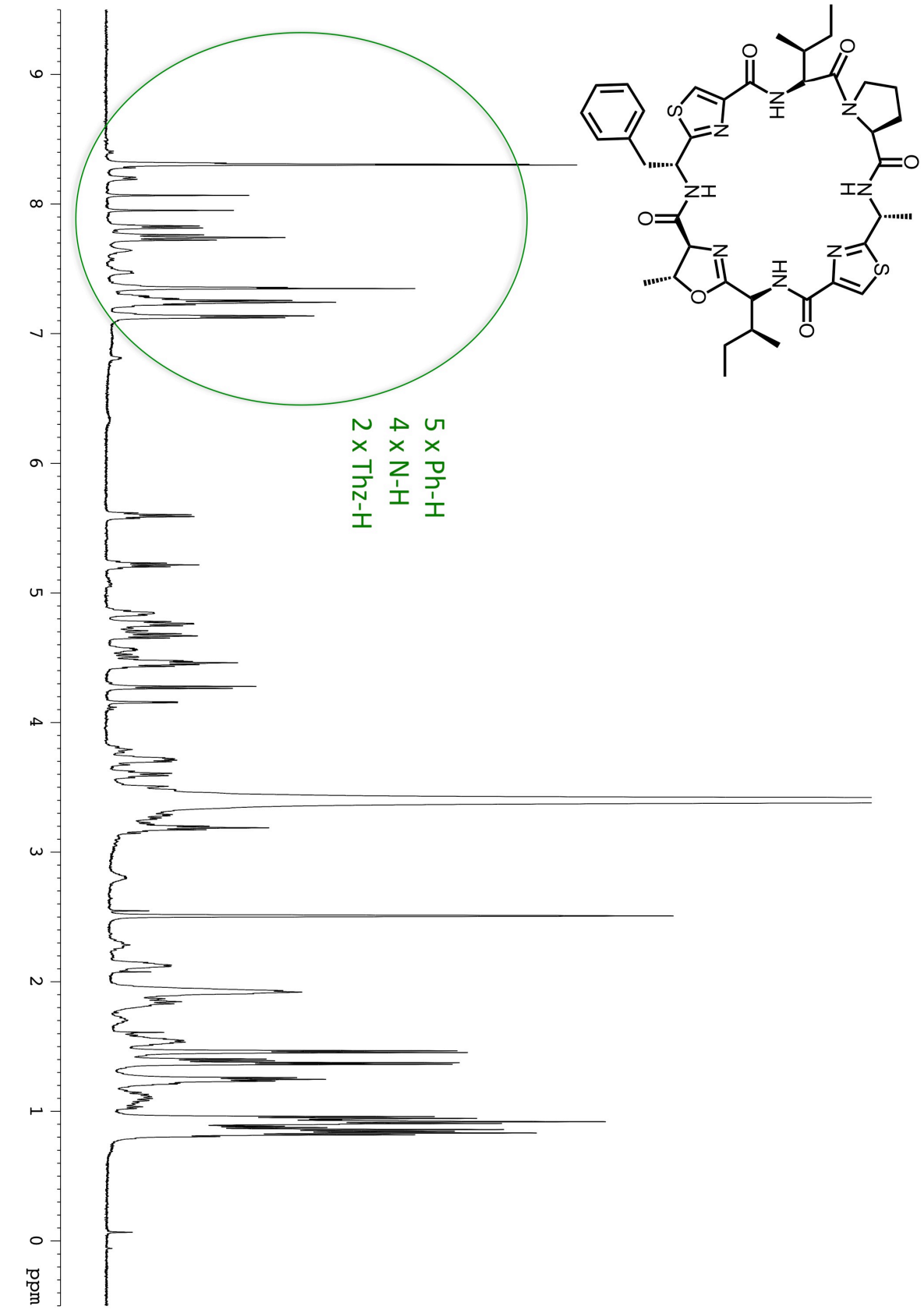



FIGURE 77: **NMR of cyclo[IPAC$^{Thz}$IT$^{MeO\times}$FC$^{Thz}$] (29)**

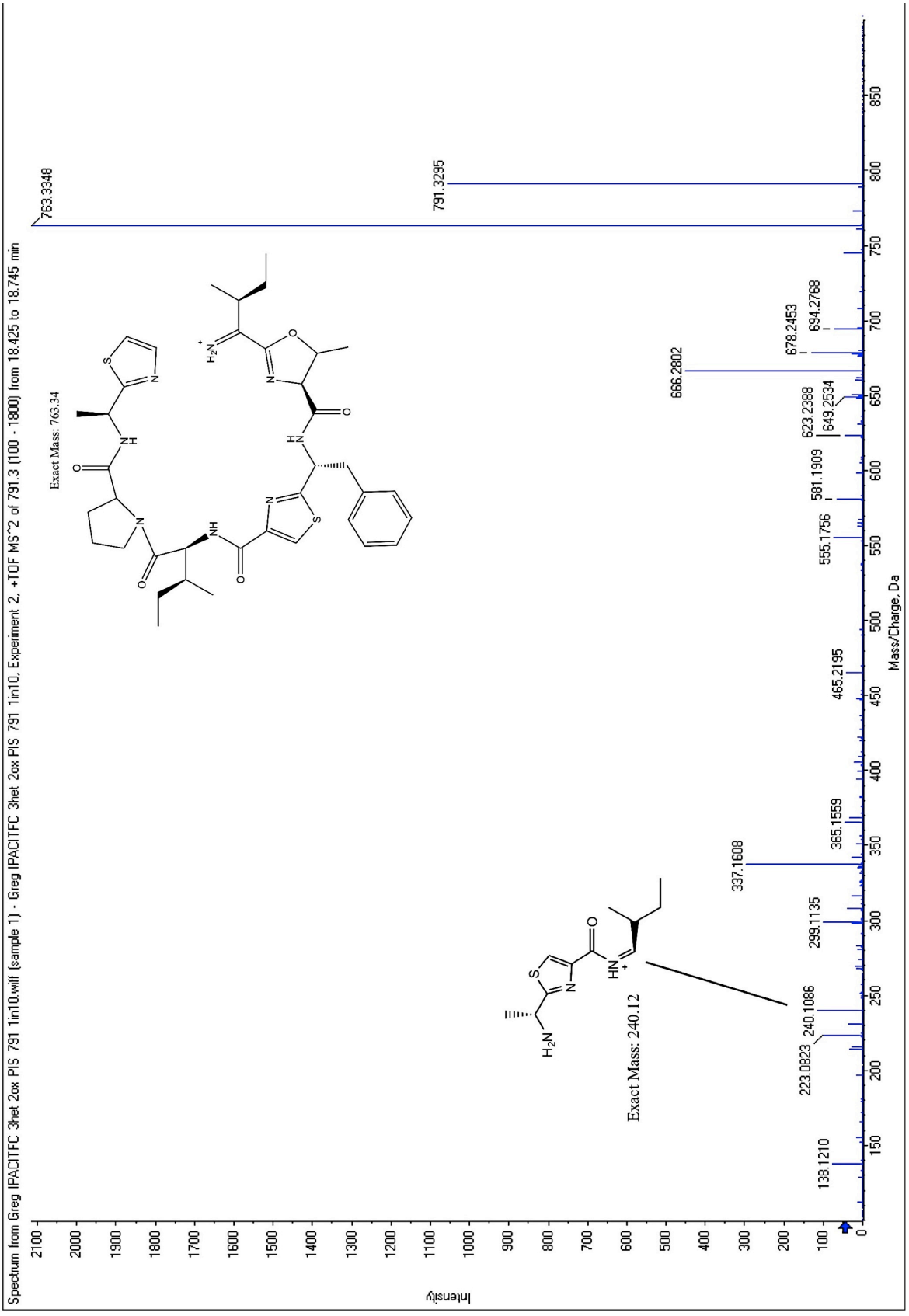



FIGURE 78: **MSMS of cyclo[IPAC$^{\mathbf{Thz}}$IT$^{\mathbf{MeOx}}$FC$^{\mathbf{Thz}}$] (29)**

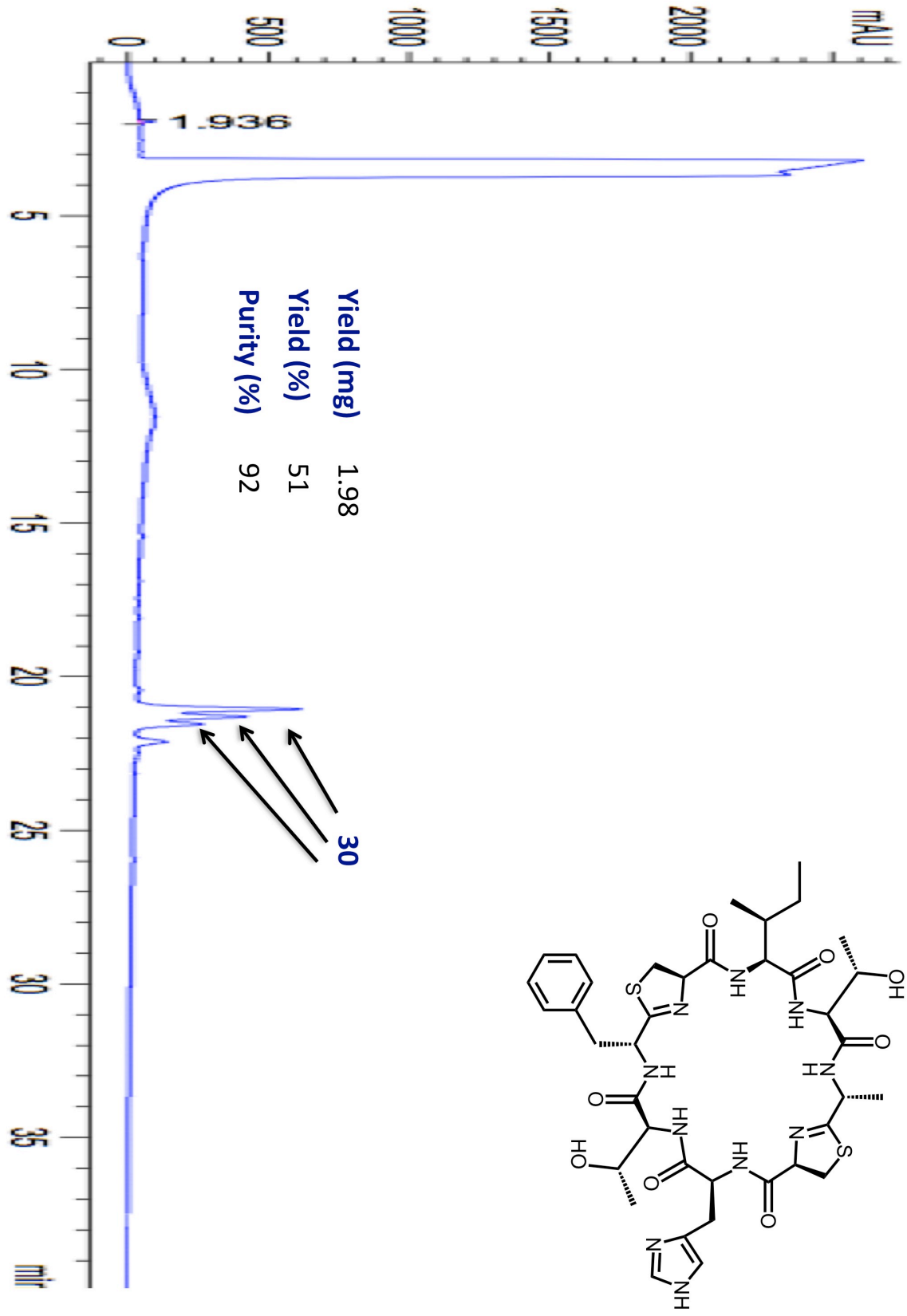



FIGURE 79: **HPLC of cyclo[ITAC$^{Tth}$HTFC$^{Tth}$] (30)**

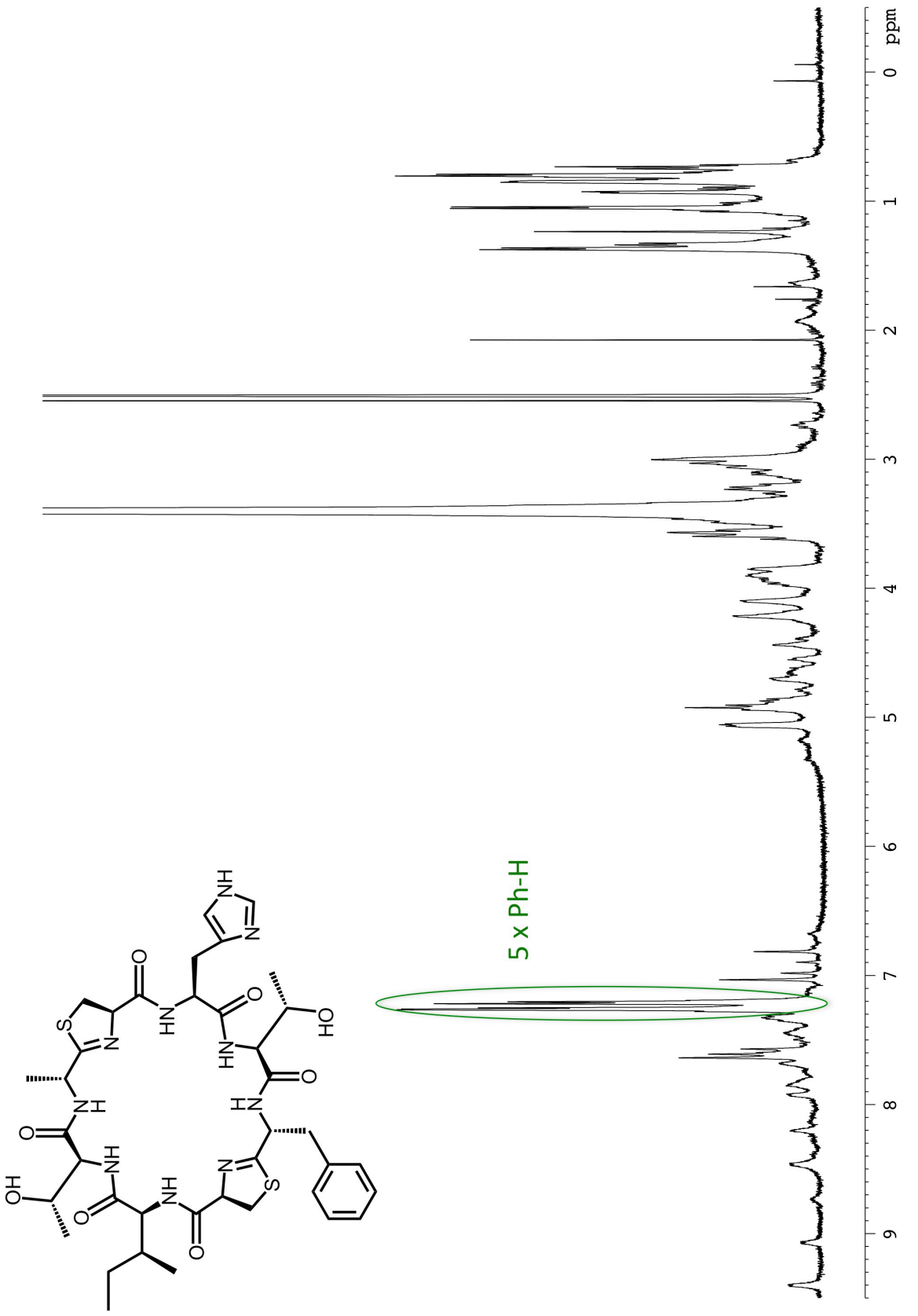



FIGURE 80: **NMR of cyclo[ITAC$^{ThH}$HTFC$^{ThH}$] (30)**

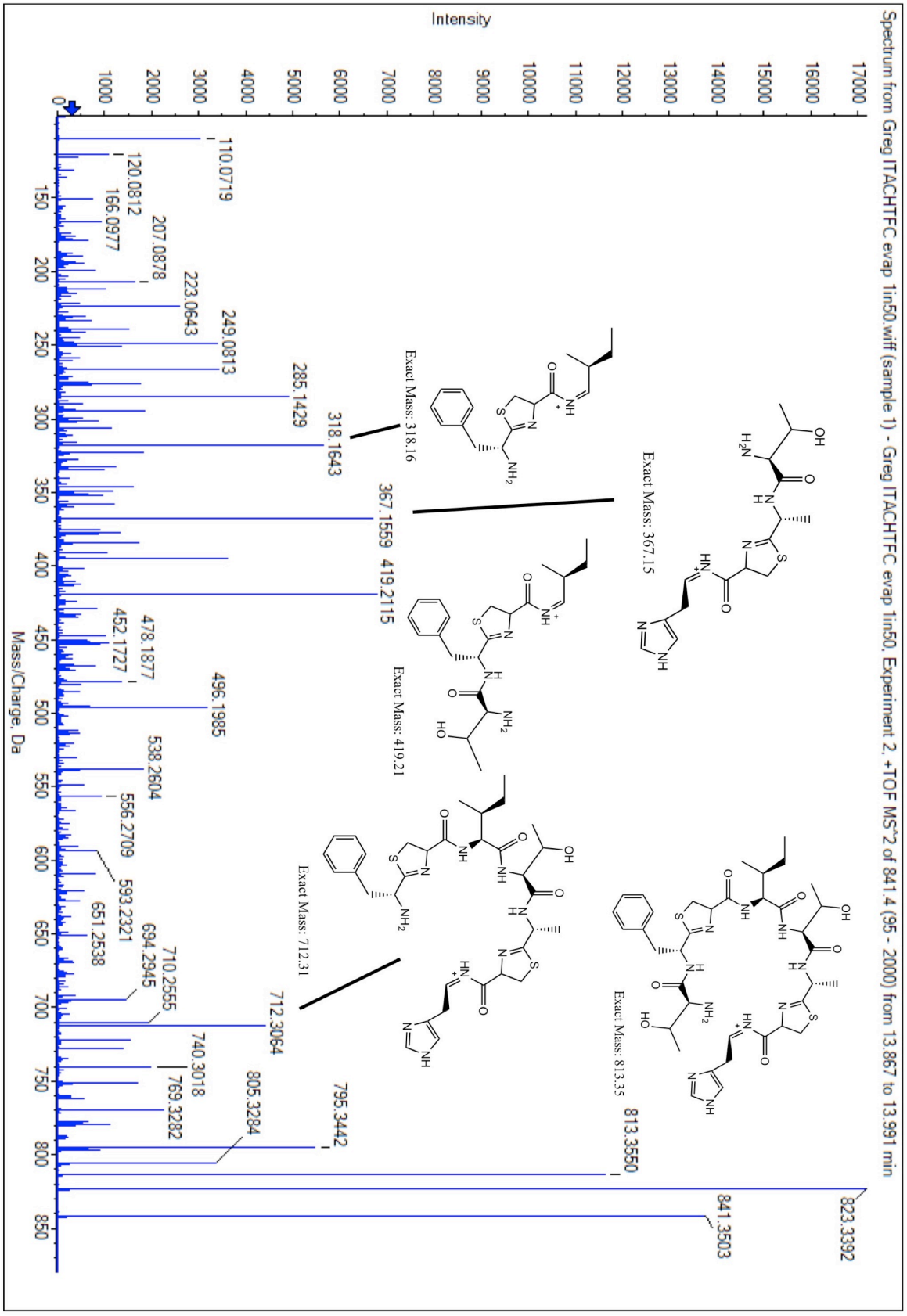



FIGURE 81: **MSMS of cyclo[ITAC$^{ThH}$HTFC$^{ThH}$] (30)**

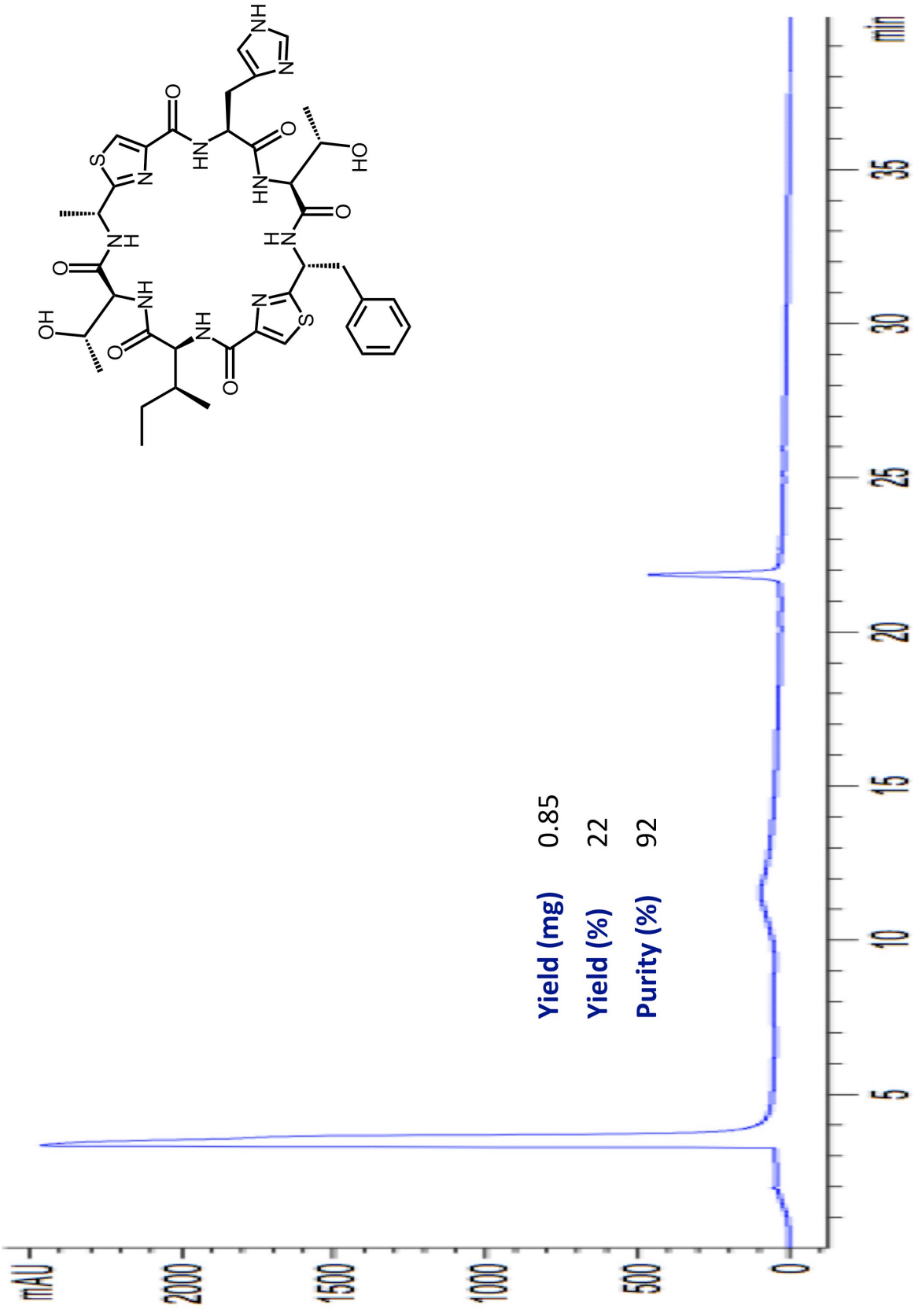



FIGURE 82: **HPLC of cyclo[ITAC$^{Thz}$HTFC$^{Thz}$] (31)**

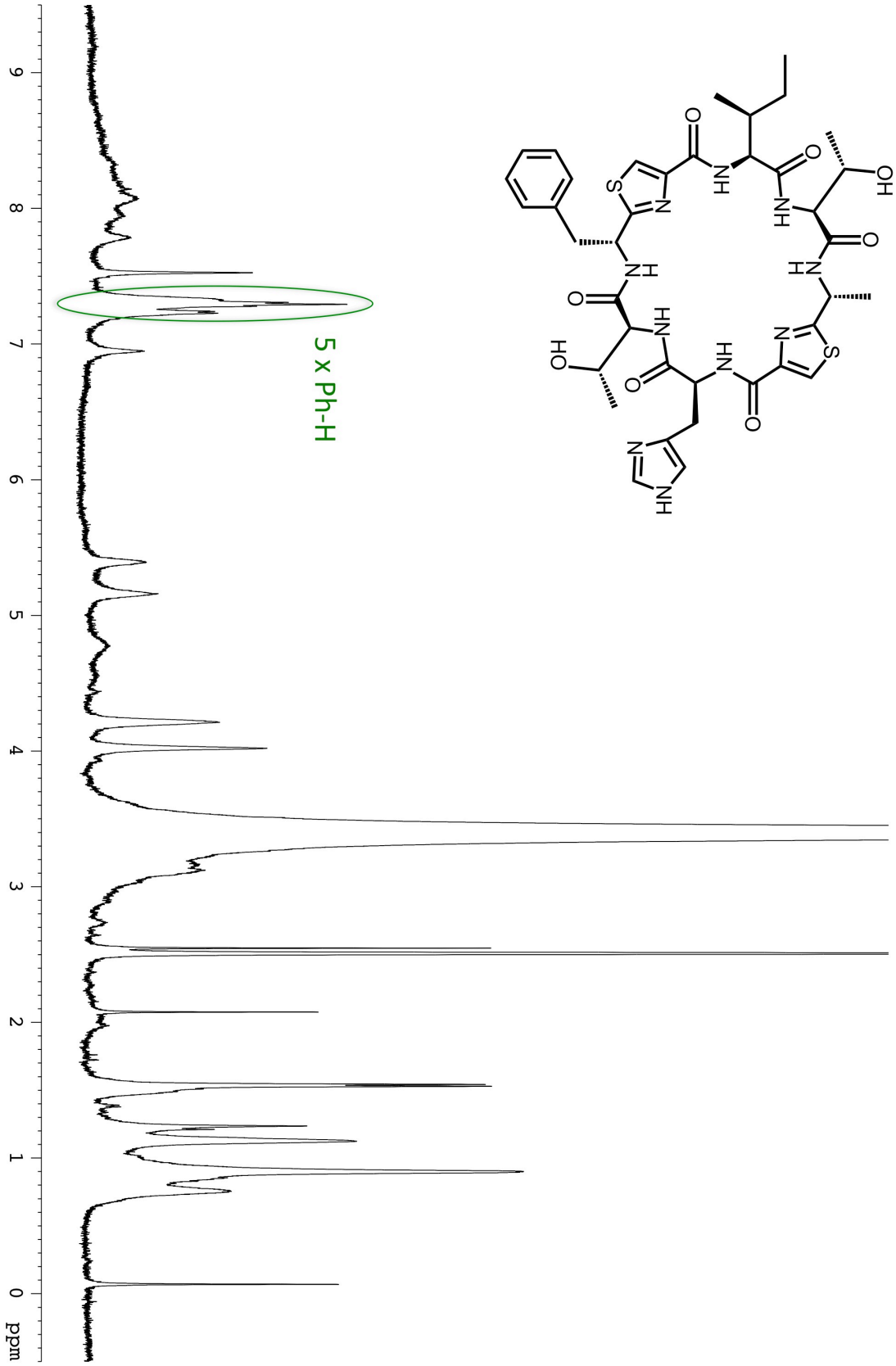



FIGURE 83: **NMR of cyclo[ITAC$^{Thz}$HTFC$^{Thz}$] (31)**

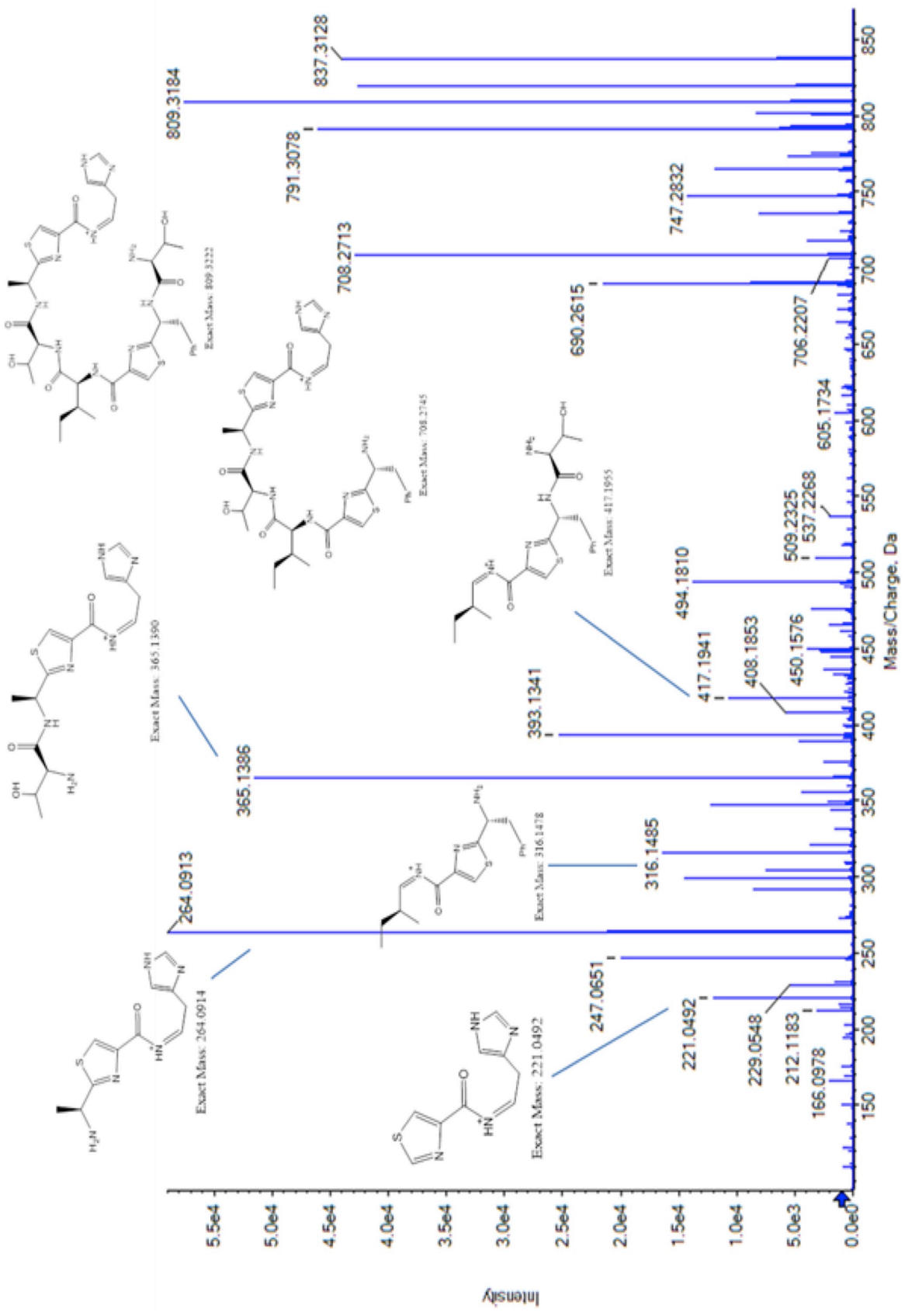



FIGURE 84: **MSMS of cyclo[ITAC$^{Thz}$HTFC$^{Thz}$] (31)**

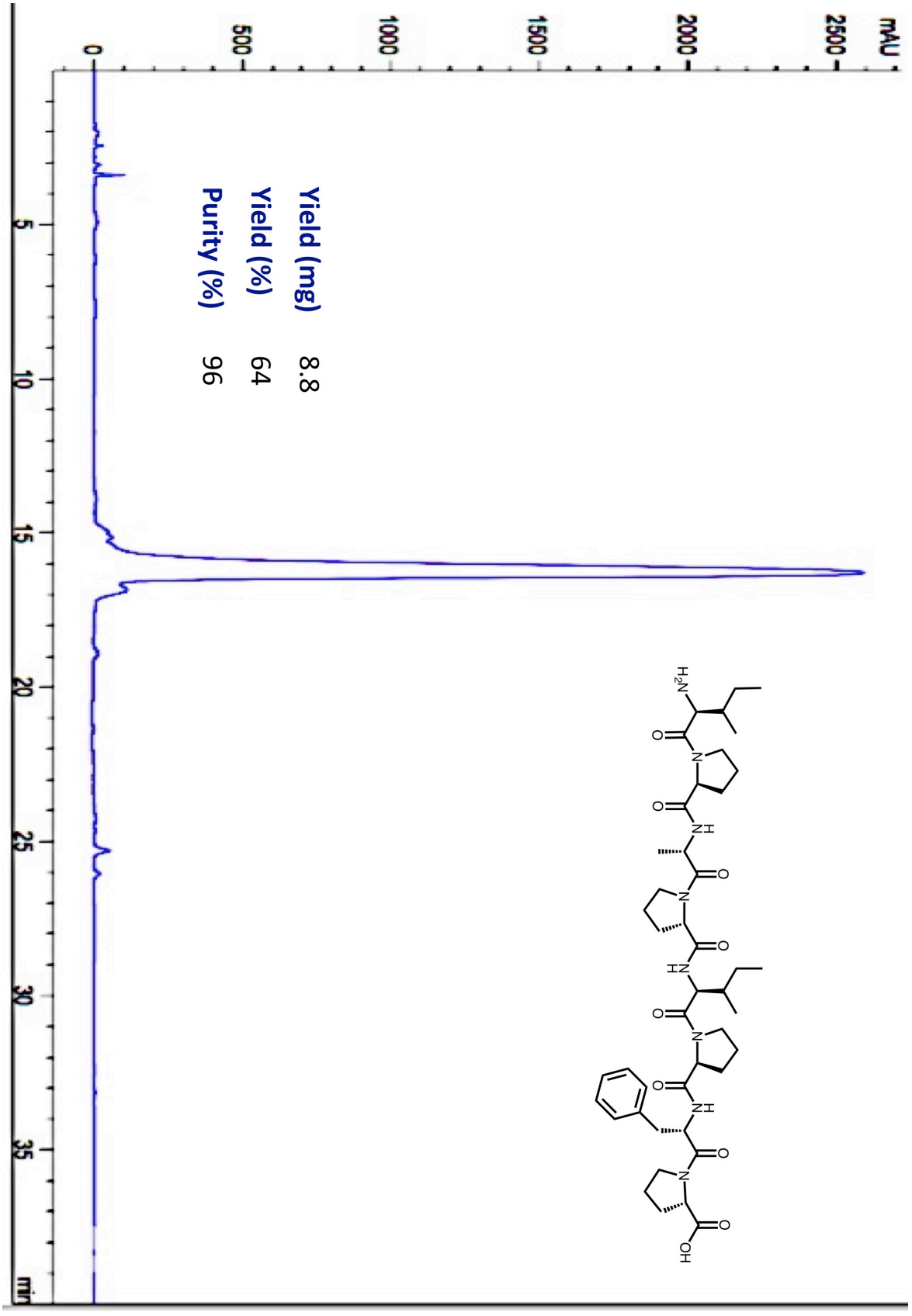



FIGURE 85: **HPLC of IPAPIPFP (32)**

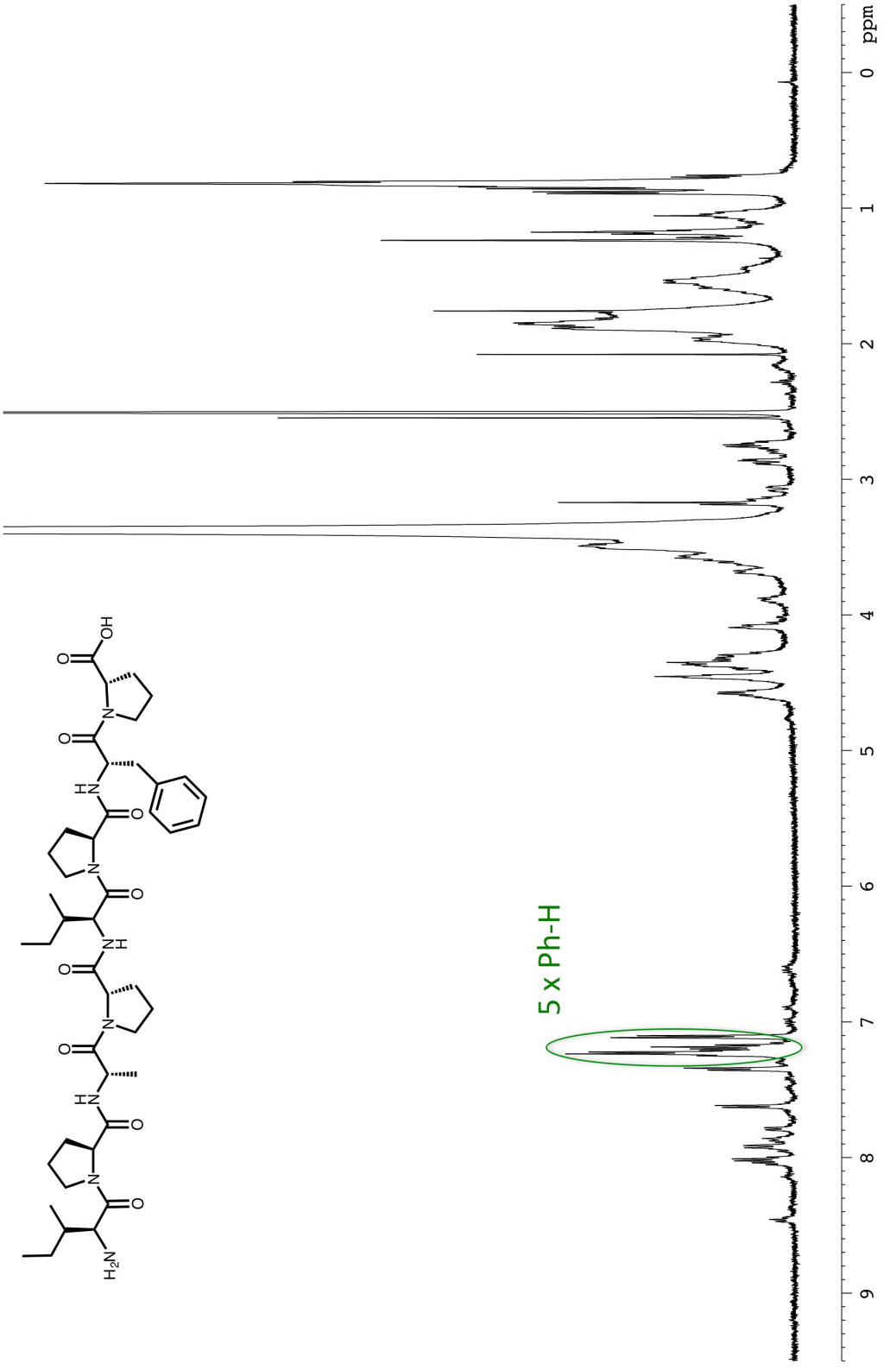



FIGURE 86: **NMR of IPAPIPFP (32)**

# Appendix C - Publications

Bent A. F.*, **Mann G.***, Houssen W. E.*, Mykhaylyk V., Duman R., Thomas L., Jaspars M., Wagner M. and Naismith J. H. Structure of the Cyanobactin Oxidase ThcOx from *Cyanothece* sp. PCC 7425, the First Structure to be Solved at Diamond Light Source Beamline I23 by means of S-SAD, *Acta Crystallograph. Sect. D Struct. Biol.*, 72, 1174 —1180 (2016) doi: 10.1107/S2059798316015850.

Oueis E., Adamnson C., **Mann G.**, Ludewig H., Redpath P., Migaud M., Westwood N. J. and Naismith J. H. Derivatisable Cyanobactin Analoues: A Semisynthetic Approach, *Chembiochem*, 16, 2646 – 2650 (2015) doi: 10.1002/cbic.201500494.

Koehnke J.*, **Mann G.***, Bent A. F.*, Ludewug H., Shirran S., Botting C., Lebl T., Houssen W. E., Jaspars M. and Naismith J. H. Structural Analysis of Leader Peptide Binding Enables Leader-free Cyanobactin Processing, *Nature Chem. Biol.*, 11, 558 – 565 (2015) doi: 10.1038/nchembio.1841.

**Mann G.**, Koehnke J., Bent A. F., Graham R., Houssen W. E., Jaspars M., Schwarz-Linek U. and Naismith J. H. The Structure of the Cyanobactin Domain of Unknown Function from PatG in the Patellamide Gene Cluster, *Acta Crystallograph. Sect. F Struct. Biol. Commun.*, 70 1597 –1603 (2014) doi: 10.1107/S2053230X1402425X.

Koehnke J.*, Bent A. F.*, Housse W. E.*, **Mann G.***, Jaspars N. and Naismith J. H. The Structural Biology of Patellamide Biosynthesis, *Curr. Opin. Struct. Biol.*, 29, 112 – 121 (2014) doi: 10.1016/j.sbi.2014.10.006

Houssen W.E., Bent A. F., McEwan A.R., Pieiller N., Tabudravu J., Koehnke J., **Mann G.**, Adaba R. I., Thomas L., Hawas U. W., Liu H., Schwarz-Linek U., Smith M. C. M., Naismith J. H. and Jaspars M. An Efficient Method for the In Vitro Production of

Azol(in)e-Based Cyclic Peptides, *Angew. Chem. Int. Ed.* 53, 14171 – 14174 (2014) doi: 10.1002/anie.201408082.

Koehnke J., Bent A. F., Zollman D., Smith K., Houssen W. E., Zu X., **Mann G.**, Lebl T., Scharff R., Shirran S., Botting C. H., Jaspars M., Schwarz-Linek U. and Naismith J. H. The Cyanobactin Heterocyclase Enzyme: A Processive Adenylase that Operates with a Defined Order of Reaction, *Angew. Chem.*, 52, 13991 – 13996 (2013) doi: 10.1002/anie.201306302.

* = Equal contribution